



Read Better Be Better

# LITERACY PROGRAM

## ANALYSIS OF PROGRAM DATA



# PREPARED FOR: READ BETTER BE BETTER

*Submitted by:*  
**Melissa Kovacs, Ph.D., PStat**  
*mkovacs@firsteval.com*  
*602.492.1406*

*Version 5*  
*March 2, 2021*

## Preface – Placement of RBBB Program Among ESSA Evidence Tiers

This report preface describes how the Read Better Be Better (“RBBB”) program is a Tier 2 Moderate Evidence level of program effectiveness. I have completed three evaluations of the RBBB program since 2016, this document reporting on the third evaluation, each of which have shown evidence of contributing to literacy gains among 3<sup>rd</sup> grade participants.

RBBB’s program evidence comes from the use of a quasi-experimental design with comparison groups that have baseline equivalence to participants. This solidly places them in Tier 2, according to evidence of effectiveness standards under ESSA.<sup>1</sup>

Specifically, Tier 2 Moderate Evidence requires:

1. A statistically significant positive program effect,
2. At least 350 students,
3. At least two educational sites, and
4. A quasi-experimental study design.

The RBBB program and study design meet all these requirements:

1. Statistically significantly higher literacy achievement was found among 3<sup>rd</sup> grade program participants as compared with their comparison group counterparts among the following literacy assessments: Reader Self-Perception Scale; Dibels ORF; Dibels Composite Score; Galileo; ORA; AIMSweb; and a teacher evaluation of the student participants. Statistical significance was determined through general linear modeling and the study’s statistical methodology is described in the Appendix to the February 2021 evaluation report.
2. Three of these assessments meet the Tier 2 sample size requirement, with the following sample size of participants: Reader Self-Perception Scale (n=1,150); Dibels ORF (n=553); and Dibels Composite Score (n=389).
3. The Reader Self-Perception Scale is implemented in 10 school districts that each have multiple schools; the Dibels ORF is implemented in 5 school districts that each have multiple schools; and the Dibels Composite Score is implemented in 4 school districts that each have multiple schools.
4. The quasi-experimental study design included collection of a large amount of comparison group data. Comparison group data consisted of student assessment data from the same school of students who were not in the RBBB program. Comparison group data was much greater than the participant group data sample sizes so as to craft a comparison group of baseline equivalence. In this sense, only comparison group students whose pre-test scores were within +/- 1.0 standard deviations of the participant group students’ pre-test scores were included in the comparison group. Pre-test and post-test scores were recorded for both participants and the comparison group. Repeated measures general linear models (“GLM”) were used to test the differences in growth between program participants and non-participants in the crafted, matched comparison groups. Data were explored for the use of propensity score matching methods but did not meet key

---

<sup>1</sup> See “Evidence-Based Interventions Under the ESSA” video at <https://www.cde.ca.gov/re/es/evidence.asp> and “Using the What Works Clearinghouse to find ESSA Tiers of Evidence” from the Institute of Education Sciences at <https://ies.ed.gov/ncee/wwc/essa>

assumptions necessary for the use of this analytical approach, resulting in the use of GLM.

Meeting the ESSA Tier 1 evidence level would require randomization of students into participating in the RBBB program, and this has not occurred. The RBBB program clearly meets all requirements of Tier 2, Moderate Evidence.

A handwritten signature in blue ink that reads "Melissa Kovacs". The signature is written in a cursive, flowing style.

Melissa Kovacs, Ph.D., PStat  
FirstEval  
February 27, 2021

## List of Tables and Figures

### List of Tables

Table 1: Description and Summary of Assessments – 3 <sup>rd</sup> Grade Analyses
Table 2: Description and Summary of Assessments – 2 <sup>nd</sup> Grade Pilot
Table 3: Description and Summary of Assessments – School Day Pilot
Table 4: RSPS Subscale Score Increases by Year
Table 5: Alignment of RBBB Data with DIBELS ORF Benchmark Goal Cut-Points
Table 6: Alignment of RBBB Data with RIT Norms

### List of Figures

Figure 1: RSPS Subscale Component Score Averages for 3rd Grade Participants
Figure 2: DIBELS DAZE Average Scores, Pre and Post Test, Fall Semesters only
Figure 3: DIBELS DAZE Average Scores, Pre and Post Test, Spring Semester only
Figure 4: Dibels DAZE Score Averages for 3rd Grade Participants; 2016 and 2020 data results shown
Figure 5: DIBELS ORF Average Scores, Pre and Post Test, Fall Semesters only
Figure 6: DIBELS ORF Average Scores, Pre and Post Test, Spring Semester only
Figure 7: Dibels ORF Score Averages for 3rd Grade Participants (n=553)
Figure 8: DIBELS Composite Score Averages, Pre and Post Test, Fall Semesters only
Figure 9: DIBELS Composite Score Averages, Pre and Post Test, Spring Semester only
Figure 10: Dibels ORF Score Averages for 3rd Grade Participants (n=777)
Figure 11:Galileo Score Averages, Pre and Post Test, Fall Semesters Only
Figure 12: Number of RBBB Students at Galileo Benchmark Goals during Fall Semesters
Figure 13: Galileo Score Averages, Pre and Post Test, Spring Semester Only
Figure 14: ORA Average Scores, Pre and Post Test
Figure 15: AIMSweb Score Gains from Pre-test to Post-test Time, Oral Reading Fluency, Reading Comprehension, and Composite Score
Figure 16: RIT Average Scores, Pre and Post Test, Fall Semesters only
Figure 17: RIT Average Scores, Pre and Post Test, Spring Semester only
Figure 18: iReady Score Gains from Pre-test to Post-test Time, Lit, Info, and Full
Figure 19: Illuminate Average Scores, Pre and Post Test
Figure 20: Number of RBBB Participants in FAME Category at Pre and Post Test Time
Figure 21: Percentage of RBBB Participants and Non-Participants Moving FAME Categories from Pre Test to Post Test Time
Figure 22: Teacher Evaluations of 3rd Grade Participants
Figure 23: Teacher Evaluations of 8th Grade Participants
Figure 24: Percentage of Leaders either Strongly Agreeing or Agreeing with each Statement, RLS
Figure 25: Summary of Data Results for 3 <sup>rd</sup> Grade Readers, by Assessment Instrument
Figure 26: RSPS Subscale Component Score Averages for 3rd Grade Participants
Figure 267: DIBELS DAZE Average Scores, Pre and Post Test, Fall Semesters only
Figure 28: DIBELS ORF Average Scores, Pre and Post Test, Fall Semesters only

Figure 29: DIBELS Composite Score Averages, Pre and Post Test, Fall Semesters only  
Figure 30: Galileo Score Averages, Pre and Post Test, Fall Semesters Only  
Figure 31: AIMSweb Scores from Pre-test to Post-test Time, 2<sup>nd</sup> Grade, n=4  
Figure 32: Number of RBBB 2<sup>nd</sup> Grade Participants in FAME Category at Pre and Post Test Time (n=15)  
Figure 33: Teacher Evaluations of 2nd Grade Participants – Unpaired Assessments  
Figure 34: Summary of Data Results for 2<sup>nd</sup> Grade Pilot Program, by Assessment  
Figure 35: RSPS Subscale Component Score Averages for 3rd Grade Participants  
Figure 36: ORA Average Scores, Pre and Post Test  
Figure 37: RIT Average Scores, Pre and Post Test, Fall Semesters only  
Figure 38: Teacher Evaluations of 3rd Grade Participants  
Figure 39: Teacher Evaluations of 6th Grade Participants  
Figure 40: Percentage of 6<sup>th</sup> Grade Leaders either Strongly Agreeing or Agreeing with each Statement, RLS

## Executive Summary

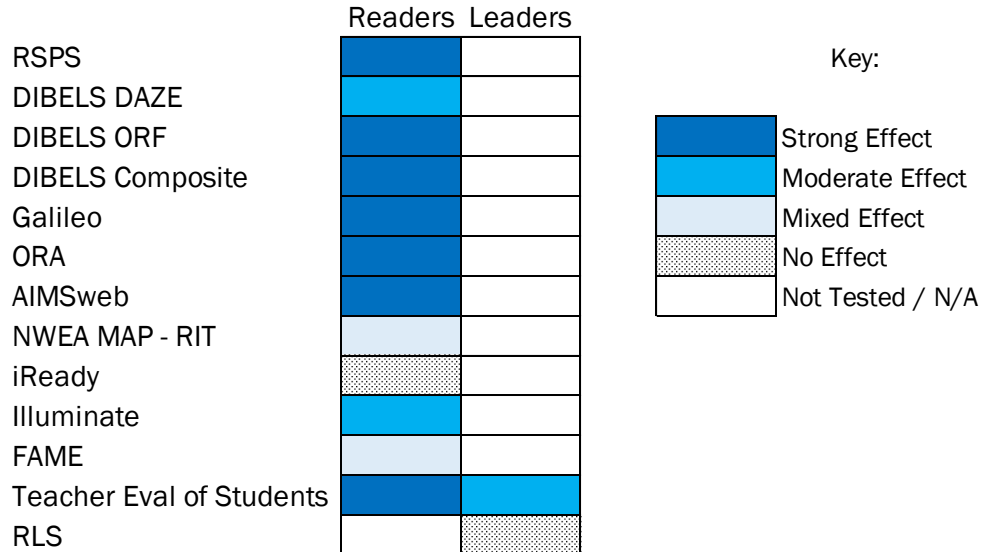
This report details results from analyzing Read Better Be Better’s program data from the Fall, 2018 through the Fall, 2019 semesters.

The Read Better Be Better (“RBBB”) program uses trained 6<sup>th</sup>, 7<sup>th</sup>, and 8<sup>th</sup> grade student volunteers (“Leaders”) to implement a structured reading program to 3<sup>rd</sup> grade students (“Readers”). The program’s mission is to help children improve their literacy skills and become better learners. The program targets Tier II students according to the Arizona State Literacy Plan.

This report represents an analysis of all available outcome data. This report is not a comprehensive evaluation report and does not address RBBB processes or program implementation. Evidence of program improvement was obtained using the following educational assessments:

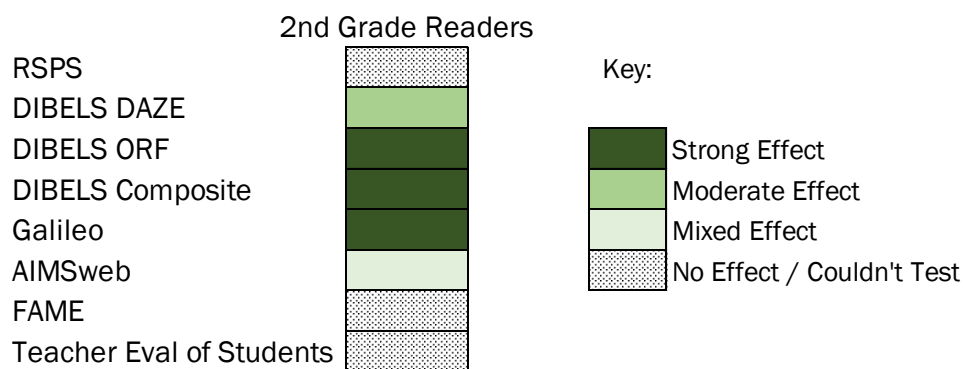
Reader Self-Perception Scale; DIBELS DAZE; ORF; and Composite; Galileo; Oral Reading Analysis; AIMSweb; NWEA MAP-RIT; iReady; Illuminate; FAME; Teacher Evaluations and Reading Leader Survey.

The RBBB program has strong positive effects on its 3<sup>rd</sup> grade participants’ literacy skills. However, the RBBB program does not show strong effects for Leaders. The figure below summarizes these findings for 3<sup>rd</sup> grade participants:



As the third evaluation report of the RBBB program, this allows for reflection on program growth and achievement. The first report, provided in 2016, showed RBBB effectiveness but had small sample sizes, as the program was still small. The 2018 report and this report benefit from larger sample sizes as more and more schools have joined the program. The 2018 report, and this one, also show RBBB effectiveness but are bolstered by increasing sample sizes as more and more schools are added to the program. This strengthens the findings that the program is having a significant impact on 3<sup>rd</sup> grade Readers and a moderate impact on 8<sup>th</sup> grade Leaders.

As well, assessments provide varying to moderate evidence that a 2<sup>nd</sup> grade pilot program for Readers impacts participants' literacy skills. While evidence for the 2<sup>nd</sup> grade pilot is positive, it is not as strong as findings for the 3<sup>rd</sup> grade Readers. The figure below summarizes the findings for the 2<sup>nd</sup> grade pilot of Readers:



For RBBB participants in the School Day pilot, there was little evidence that their program participation was impactful. This may be due to a small sample size.

As well, little evidence was found that a matched Leader contributes to Reader achievement.



## Introduction

This report presents results from analyzing Read Better Be Better's program data for three semesters: Fall, 2018 through Fall, 2019.

In January of 2015, Read Better Be Better ("RBBB") piloted an after-school literacy program that pairs 8<sup>th</sup> grade students to help 3<sup>rd</sup> grade students become better readers. The program has grown steadily since the pilot and is now operating in ten school districts. The RBBB program uses trained 6<sup>th</sup>, 7<sup>th</sup>, and 8<sup>th</sup> grade student volunteers to implement a structured reading program to 3<sup>rd</sup> grade students. The older students ("Leaders") work one-on-one with paired younger students ("Readers") to model comprehension strategies, help with reading skills, and provide activities that improve focus and concentration.

This is the third report of RBBB program data results, after reports provided in 2016 and 2018.

RBBB's mission and vision are detailed below, and the RBBB Logic Model is included as Appendix A to this report.

The Mission: Read Better Be Better connects young readers and youth leaders to inspire a love of literacy and learning.

The Vision: A society in which all children master the foundational skills necessary to become independent learners.

The RBBB program is intended to target Tier II students, as defined by the Arizona State Literacy Plan. In the Arizona State Literacy Plan, Tier I students are in need of "universal instruction," consisting of a core reading program and benchmark testing. Tier II students need an additional small group intervention beyond Tier I instruction. Tier III students need intensive instruction and remediation services.<sup>2</sup>

Commonly, literacy programs address Tier III students, while RBBB specifically targets Tier II students. Tier II students are often the ones "falling through the cracks," in that they do need additional literacy help, but are not the students in their schools who struggle the most with reading. Therefore, when resources are scarce, Tier II students' needs often remain minimally addressed at best.

Different school districts use different assessments to measure reading improvement. This report is organized by assessment, and each section details which school district uses which assessment.

---

<sup>2</sup> The Arizona State Literacy Plan can be found here: <http://www.azpromisingpractices.com/LiteracyforRTI.pdf> and the definitions of Tier I, II, and III are found on pages 6-7.

## RBBB in Context

RBBB's cross-age peer tutoring model comes from a strong foundation of proven effectiveness. Numerous studies find that cross-age peer tutoring is beneficial for both the younger and older grade participating students.<sup>3</sup> Some studies also found that participation in these programs boosts views of oneself as a reader and leads to less negative thinking about reading.<sup>4</sup> A study in Syracuse found that tutees' participation in a cross-age peer reading program engendered bigger gains than their tutors experienced.<sup>5</sup>

One researcher found that the effects of participating in a cross-age peer tutoring reading program are stronger in later phases of the program.<sup>6</sup> This is important evidence for RBBB to monitor its long-term outcomes. RBBB's outcomes are displayed in the logic model in Appendix A.

## Scope of This Report

This report represents an analysis of outcome data for all existing RBBB program data, inclusively covering Fall, 2018 through Fall, 2019. This report is not a comprehensive evaluation report. FirstEval did not evaluate RBBB processes or program implementation. FirstEval did, however, analyze all existing RBBB program participant and comparison group data provided by RBBB. In this sense, this report addresses most of the short-term outcomes in the RBBB logic model, and all of the medium-term outcomes in the logic model. This report does not address any of the long-term outcomes in the logic model. The RBBB logic model, highlighted to show this report's scope, is attached as Appendix A.

## Structure of This Report

The primary focus of this report is the 3<sup>rd</sup> grade Readers and their literacy progress with their 6<sup>th</sup> / 7<sup>th</sup> / 8<sup>th</sup> grade Leaders. This report also describes two pilot programs implemented by RBBB since the last evaluation report – a pilot of 2<sup>nd</sup> grade students, and a pilot program of implementing the RBBB program during the school day, instead of after school. At a high level, this report is structured around three sections for each of these populations (traditional 3<sup>rd</sup> grade

---

<sup>3</sup> See, for example, Loretta Abassi, Cleveland State University, "Effects of Cross-Age Tutoring on Reading Attitudes of Elementary School Students;" John Hattie, 2006, "Cross-Age Tutoring and the Reading Together program," in *Studies in Educational Evaluation*; Van Keer et al., 2005, "Effects of Explicit Reading Strategies Instruction and Peer Tutoring on Second and Fifth Graders' Reading Comprehension and Self-Efficacy Perceptions," in the *Journal of Experimental Education*; Wright and Cleary, 2006, "Kids in the Tutor Seat: Building Schools' Capacity to Help Struggling Readers Through a Cross-Age Peer-Tutoring Program," in *Psychology in the Schools*; and Slavin and Madden, 1989, "What Works for Students at Risk: A Research Synthesis," in *Educational Leadership*.

<sup>4</sup> See Abassi and Van Keer

<sup>5</sup> See Wright and Cleary

<sup>6</sup> See Hattie

Readers and 6<sup>th</sup> / 7<sup>th</sup> / 8<sup>th</sup> grade Leaders; the 2<sup>nd</sup> grade pilot Readers; and the School Day pilot) and then assessment results are discussed within each of these population sections.

## Assessment Instruments

Data from numerous educational assessment instruments were provided by RBBB to FirstEval for analysis. This section describes those instruments and how this report is structured around the instruments.

This report examines data from Readers and Leaders separately and examines data among the different assessment instruments separately. RBBB has non-program participant data from numerous assessments. Non-participant data provide an ideal group with which to construct a matched comparison group to assess the impact of the RBBB program on reading progress.

Because RBBB program participation was not randomly assigned, these non-participants do not represent a true control group, but rather a matched comparison group. More on the statistical methodology of comparing participants' progress with comparison group students' progress is in Appendix F.

Next, each assessment instrument is described in the order its results will be reported in the following section.

~~

### *Reader Self-Perception Scale*

The Reader Self-Perception Scale ("RSPS") is RBBB's primary reading efficacy tool, and is a tool to measure how children feel about themselves as readers.<sup>7</sup> The instrument consists of 33 items, and is divided into 5 subscales for analytic purposes. The subscales measure general perception of one's own reading, progress (how one's perception of present reading performance compares with past performance), observational comparison (how a child perceives her or his reading performance to compare with the performance of classmates), social feedback (direct or indirect input about reading from teachers, classmates, and people in the child's family), and physiological state (internal feelings that the child experiences during reading). These subscales have been shown to have high internal consistency and reliabilities.<sup>8</sup> The instrument is included as Appendix B.

The RSPS fits neatly into measuring progress towards RBBB's mission and vision, by measuring an active enjoyment of reading.

---

<sup>7</sup> See Henk & Melnick, 1995. "The Reader Self-Perception Scale (RSPS): A New Tool for Measuring How Children Feel About Themselves as Readers," in *The Reading Teacher*, Vol. 48 No. 6.

<sup>8</sup> See Henk & Melnick mentioned previously.

### *Dibels*

RBBB receives data from participating schools on students' Dibels scores. Dibels is an acronym for Dynamic Indicators of Basic Early Literacy Skills. The Dibels family of instruments is widely used and was developed at the University of Oregon. For purposes of this report, the Dibels DAZE and the Dibels ORF data provided useful information to gauge RBBB participants' progress.

### *Dibels DAZE*

DAZE, or the DIBELS maze comprehension task, is a group-administered measure of reading comprehension. According to the University of Oregon website, students are asked to read a passage silently. In the passage, every seventh word (approximately) is blank, with a maze of options (i.e., three possible word choices for the blank). One of the words in the maze is always correct, and the other two are incorrect. DAZE requires students to choose the correct word as they read the passage. Students are given three minutes to work on this task, and the results are scored.

The instrument has shown strong reliability and validity, and more information can be found here: <https://dibels.uoregon.edu/assessment/dibels/measures/daze.php>

### *Dibels ORF*

Dibels ORF (oral reading fluency) is another comprehension tool that is individually administered to test for reading fluency. Students read passages while an administrator records miscues, then the student retells the passage, hitting certain highlights in the passage to prove comprehension.

More information about the Dibels ORF instrument can be found here: <https://dibels.uoregon.edu/assessment/dibels/measures/orf.php>

### *Galileo*

Galileo began at the University of Arizona in 1986 to examine whether students are ready; almost ready; or ready later to learn literacy concepts. Currently it integrates district curriculum with standards (like Common Core) and comprehension assessments. The assessments can be taken on computer, paper, or hand-held devices. Galileo exams are benchmark exams, intended to inform teachers how students are progressing through their school year.

The Galileo instrument is proprietary, and therefore not attached as an Appendix to this report. More information about Galileo can be found here: <http://www.ati-online.com/>

### *Oral Reading Analysis*

RBBB receives data from the Osborn school district on students' Oral Reading Analysis ("ORA") scores. The ORA scores stem from two assessments - Fountas & Pinnell Benchmark Assessment System and Rigby's Reading Assessment. These are individual assessments of students, and data was provided with goal benchmarks at the student level.

More information about the Fountas & Pinnell Benchmark Assessment System can be found here: <http://www.fountasandpinnell.com/bas/>

### *AIMSweb*

AIMSweb is a national, computer-based test to guide response to intervention and help place students in reading and math groups. AIMS is owned by the Pearson company, and more information can be found here: <http://www.aimsweb.com/>

AIMSweb is not the same as Arizona's Instrument to Measure Success, or "AIMS."

### *NWEA MAP Test – RIT*

The Northwest Evaluation Association ("NWEA") provides MAP interim assessments for reading measured with a label called "RIT." "RIT stands for Rasch Unit and is a measurement scale developed to simplify the interpretation of test scores. It is an equal-interval scale, like feet and inches on a ruler, so scores can be added together to calculate accurate class or school averages. RIT scores range from about 100–300. Students typically start at the 180–200 level in the third grade and progress to the 220–260 level by high school. RIT scores make it possible to follow a student's educational growth from year to year."<sup>9</sup>

### *iReady*

Alhambra School District uses the iReady assessments. More information about iReady can be found here: <https://www.curriculumassociates.com/products/i-ready/i-ready-assessment>

### *Illuminate*

The Littleton and Tolleson school districts employ the Illuminate test to assess reading. More information about the Illuminate assessment can be found here: <https://www.fastbridge.org/>

---

<sup>9</sup> See "Six Commonly Used MAP Growth Terms Worth Knowing" by Joi Converse.  
<https://www.nwea.org/blog/2016/six-commonly-used-map-growth-terms-worth-knowing/>

## FAME

The FAME assessment is employed in the Avondale school district, and is measured in benchmark categories of Minimally Proficient, Partially Proficient, Proficient, and Highly Proficient.

### *Teacher Evaluation of Students – “Reader Teacher Survey”*

Teachers are asked to answer four brief questions of RBBB program students to assess whether the student is able to focus on the reading material; whether the student appears to enjoy reading; whether the student understands reading material in class; and whether they attend class on a regular basis. This data is examined for changes between the beginning of the semester and the end of the semester, and these questions are included as Appendix C.

### *Teacher Evaluation of Leaders – “Leader Teacher Survey”*

Teachers are also asked to answer seven brief questions about the RBBB Leaders to assess their participation in the program. This data is examined for changes between the beginning of the semester and the end of the semester, and these questions are included as Appendix D.

### *Reading Leader Survey*

RBBB surveys its Leaders to assess their progress from participating in the program. Beginning in the Spring, 2019 semester, a new Reading Leader Survey was implemented. Data are available for the following districts: Alhambra, Avondale, BGC, Buckeye, Glendale, Littleton, Osborn, Phoenix, Tolleson, and Union. The instrument is attached as Appendix E. The instrument assesses concepts such as leadership, self-confidence, and school enjoyment.

## Methodology

RBBB provided clean data to FirstEval to test for progress among program participants and non-participants, as measured by the instruments discussed in the previous section. RBBB receives the data itself from the participating schools. Data are provided at the individual student level.

FirstEval compiled and analyzed data to test for differences between baseline scores and post-program scores among participants. Analyses also tested for improvement among the comparison groups and compared improvement rates between participant groups and comparison groups. When sample sizes allowed, and when appropriate, paired-samples t-tests and repeated measures general linear models were employed.<sup>10</sup>

---

<sup>10</sup> Paired samples t-tests were used to compare student growth from pre-test to post-test time. Repeated measures general linear models were used to test the differences in growth between program participants

As well, many assessments have published benchmark scores to gauge how well local students are doing compared with national-level progress. These benchmarks are published data points that show the cut points for risk among students. When applicable, these benchmarks are visualized in this report along with the achievement results in the assessment's reporting section of this report. Comparing program effects to benchmarks helps account for outside factors and isolate the RBBB program effect. The following tables summarize the assessments and where they were employed for the RBBB program and each of the two pilot programs.

***Table 1: Description and Summary of Assessments – 3<sup>rd</sup> Grade Analyses***

<u>Assessment</u>	<u>Districts</u>	<u>Benchmarks?</u>	<u>Non-Participants?</u>
RSPS	Alhambra, Avondale, BGC, Buckeye, Glendale, Littleton, Osborn, Phoenix, Tolleson, Union	no	no
DIBELS Daze	Avondale, Buckeye, Littleton, Union.	yes	yes
DIBELS ORF	Avondale, Buckeye, Littleton, Glendale, Union.	yes	yes
DIBELS Composite	Buckeye, Littleton, Glendale, Union.	yes	yes
Galileo	Avondale	yes	yes
ORA	Osborn	no	yes
AIMSweb	Phoenix	no	yes
NWEA MAP - RIT	Osborn	yes	yes
iReady	Alhambra	no	yes
Illuminate	Tolleson, Littleton	no	yes
FAME	Avondale	yes	yes
Teacher Evaluation of Students	Alhambra, Avondale, Buckeye, Littleton, Osborn, Phoenix, Tolleson, Union	no	no
RLS	Alhambra, Avondale, BGC, Buckeye, Glendale, Littleton, Osborn, Phoenix, Tolleson, Union	no	no

---

and non-participants in crafted, matched comparison groups. See Appendix F for more on the methodology.

***Table 2: Description and Summary of Assessments – 2<sup>nd</sup> Grade Pilot***

<u>Assessment</u>	<u>Districts</u>	<u>Benchmarks?</u>	<u>Non-Participants?</u>
RSPS	Avondale, Glendale, Phoenix	no	no
DIBELS Daze	Avondale, Glendale, Phoenix	no	yes
DIBELS ORF	Avondale, Glendale	yes	yes
DIBELS Composite	Glendale	yes	yes
Galileo	Avondale	no	yes
AIMSweb	Phoenix	no	no
FAME	Avondale	no	no
Teacher Evaluation of Students	Avondale, Glendale, Phoenix	no	no

***Table 3: Description and Summary of Assessments – School Day Pilot***

<u>Assessment</u>	<u>Districts</u>	<u>Benchmarks?</u>	<u>Non-Participants?</u>
RSPS	Osborn	no	no
ORA	Osborn	no	yes
NWEA MAP - RIT	Osborn	no	yes
Teacher Evaluation of Students	Osborn	no	no
RLS	Osborn	no	no

## **Results – RBBB Program – 3<sup>rd</sup> Grade Readers and Leaders**

This section discusses results related to the primary RBBB program of 3<sup>rd</sup> grade Readers and 6<sup>th</sup> / 7<sup>th</sup> / 8<sup>th</sup> grade Leaders. Two additional pilot programs are discussed separately below.

### *Reader Self-Perception Scale (“RSPS”)*

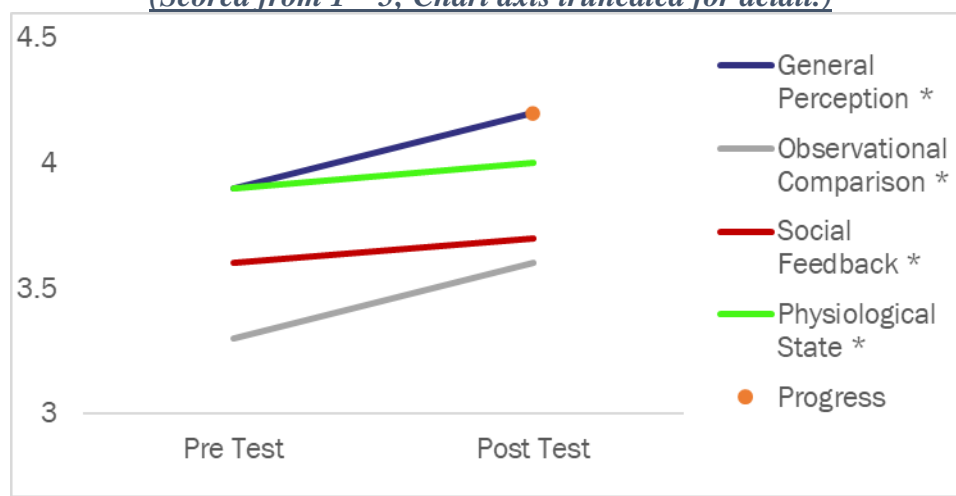
The RSPS was implemented throughout all semesters in this report’s time period (Fall 2018 – Fall 2019) and was conducted in the following districts: Alhambra, Avondale, BGC, Buckeye, Glendale, Littleton, Osborn, Phoenix, Tolleson, and Union.

Figure 1 below shows results of 3<sup>rd</sup> grade participants’ changes in scores on the RSPS. The RSPS subscales denoted here are scored between one and five, with a higher score indicating greater achievement. The Progress subscale was only scored at post-test time, since its purpose is to quantify how one’s perception of present reading performance compares with past performance. For all other subscales (Social Feedback, Physiological State, General Perception, and



Observational Comparison), RBBB participating students demonstrated statistically significant improvement in their scores between pre-test time and post-test time.

***Figure 1: RSPS Subscale Component Score Averages for 3rd Grade Participants  
(Scored from 1 – 5; Chart axis truncated for detail.)***



	Pre Test	Post Test	Score Increase
General Perception	3.9	4.2	0.3 *
Progress	N/A	4.2	N/A
Observational Comparison	3.3	3.6	0.3 *
Social Feedback	3.6	3.7	0.1 *
Physiological State	3.9	4.0	0.1 *

*\* Denotes a statistically significant improvement from pre- to post-time measures. Paired samples t-tests were conducted with significance set at the  $\alpha = .05$  level.*

This analysis also explored the potential impact that attendance in RBBB sessions could have on progress, when data was available. For the RSPS, on average, the number of RBBB program sessions attended was positively correlated with greater increases on the RSPS subscale scores. Specifically, the more RBBB program sessions a student attended, the higher the student increased their own rated Observational Comparison, Social Feedback, and Physiological State scores. The correlation between number of sessions attended and the General Perception score nears zero. These correlations are not statistically significant.<sup>11</sup>

Some changes in subscale improvement among RBBB program participants by year are noted in the RSPS instrument. Table 4 below shows the differences in subscale score increases by year for the RSPS instrument:

<sup>11</sup> Bivariate Pearson correlation with statistical significance measured at the  $\alpha = .05$  level.

***Table 4: RSPS Subscale Score Increases by Year***

	2015-16 Score Increase	2016-17 Score Increase	2017-18 Score Increase	2018-19 Score Increase
General Perception	0.49 *	0.25 *	0.17 *	0.30*
Observational Comparison	0.25 *	0.24 *	0.08	0.20*
Social Feedback	0.13	0.07	0.07	0.20*
Physiological State	0.16 *	0.06	0.10	0.15*

*\* Denotes a statistically significant improvement from pre- to post-time measures, at the  $\alpha = .05$  level, according to paired samples t-tests.*

### ***DIBELS DAZE***

The DIBELS DAZE was implemented throughout all semesters in this report's time period (Fall 2018 – Fall 2019) and was conducted in the following districts: Avondale, Buckeye, Littleton, and Union. DAZE data include both RBBB program participants and non-participants' scores. These scores are compared with published DAZE benchmark goal scores.<sup>12</sup> In order to align with the benchmark goal scores, results are reported separately for all fall semesters and the spring semester. The same middle-of-year benchmark score was used twice to compare against the fall semesters' post-test and the spring semester's pre-test. The following table describes this alignment:

***Table 5: Alignment of RBBB Data with DIBELS ORF Benchmark Goal Cut-Points***

<b>RBBB Data Time</b>	<b>DIBELS ORF Benchmark Time</b>
Fall Semester, Pre-Test	Beginning of Year
Fall Semester, Post-Test	Middle of Year
Spring Semester, Pre-Test	Middle of Year
Spring Semester, Post-Test	End of Year

This analysis applies a matched-comparison group approach when analyzing the effect of RBBB program participation on DAZE scores. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean DAZE score of RBBB program participants, within one standard deviation of the participants' mean score.<sup>13</sup>

<sup>12</sup> All Dibels benchmark scores can be found here:

<https://www.nncsk12.org/site/handlers/filedownload.ashx?moduleinstanceid=6044&dataid=5471&FileName=DIBELSNextBenchmarkGoals-7.pdf>

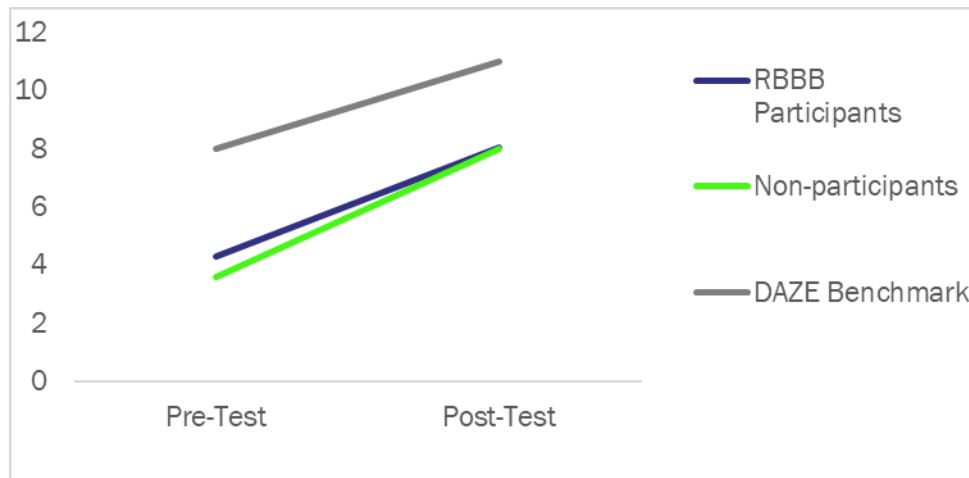
<sup>13</sup> A matched comparison-group approach follows the lead of this white paper: "Matched Comparison Group Design: An Evaluation Brief for Educational Stakeholders." Jan. 2017, Hanita et al.

[https://www.edc.org/sites/default/files/uploads/matched\\_comparison\\_group\\_design.pdf](https://www.edc.org/sites/default/files/uploads/matched_comparison_group_design.pdf)

Fall semesters only:

For Readers in the RBBB program, during both Fall semesters (2018 and 2019), their DIBELS DAZE scores increased 3.75 points, on average. The matched comparison group's score increased 4.4 points on average, while the benchmark goal shows a 3-point score increase. The following visualization and table show these results.

**Figure 2: DIBELS DAZE Average Scores, Pre and Post Test, Fall Semesters only**



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	4.30	8.05	3.75	n=468
Non-participants	3.61	8.0	4.39	n=1,569
DAZE Benchmark Goal	8	11	3	

RBBB participants started at pre-test time with a higher average DAZE score than their matched comparison group. And, RBBB participants' score gains weren't as high as the comparison group's gains. However, RBBB participants gained more than the benchmark goal recommendations.

It is important to mention that the comparison group of non-participants is an imperfect comparison group. It is much larger than the participant group, and sample sizes can impact results when comparing averages, as shown in the table above. As well, the participants and the non-participant comparison group are similarly matched, but not perfectly matched, with respect to free and reduced lunch status and ethnic / racial make-up.

Readers, Spring semester only:

For 3<sup>rd</sup> grade Readers in the RBBB program, during the Spring 2019 semester, their DIBELS DAZE scores increased 6.1 points, on average. The comparison group's score increased 6.3

points on average, while the benchmark goal recommends an 8-point score increase. The following visualization and table show these results.

***Figure 3: DIBELS DAZE Average Scores, Pre and Post Test, Spring Semester only***



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	8.21	14.31	6.1	n=253
Non-participants	7.99	14.33	6.34	n=811
DAZE Benchmark Goal	11	19	8	

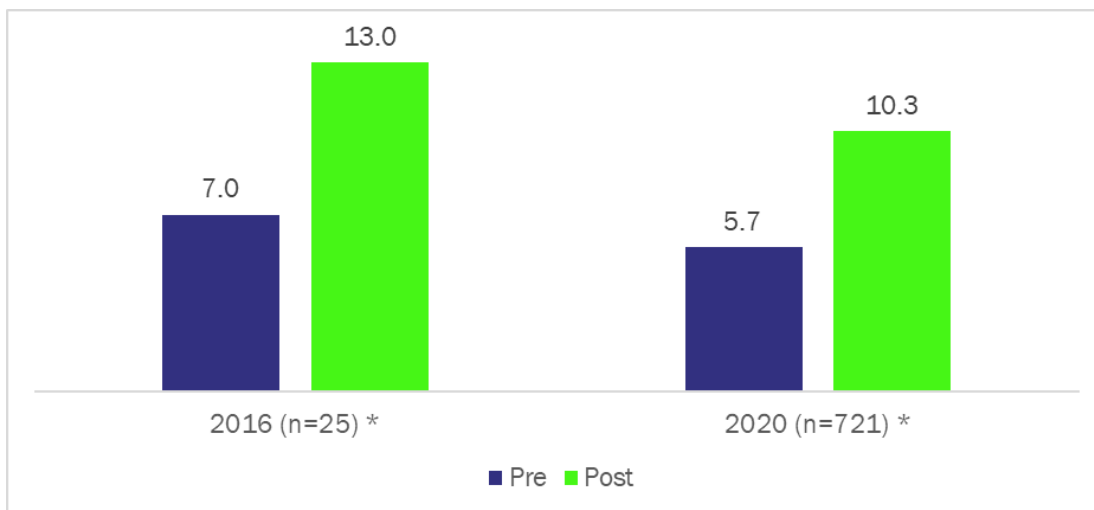
Again, RBBB participants started at pre-test time with a slightly higher average score than their comparison group counterparts. Their gains were almost equal to the comparison group's gains. And, again, the size of the comparison group is much larger than the participant group. The participants and the non-participant comparison group are similar with respect to free and reduced lunch status, while the RBBB participant group has more African-American students than the comparison group during this semester. Also, neither group meets the DAZE benchmark goal scores or the DAZE benchmark goal score increase.

Even though the DAZE scores for program participants do not rise as much as the DAZE scores for non-participants during the examined semesters, there is still a large gain in scores for the RBBB participant students.

Another way of looking at the DAZE data is to combine all the semesters and examine the RBBB participants' score increases over time. Figure 4 below shows results of 3<sup>rd</sup> grade participants' changes in average DAZE scores. RBBB participating students demonstrate statistically significant improvement in their scores between pre-test time and post-test time. The figure also shows the DAZE scores from the 2016 evaluation report. In 2016, there was very little data on the DAZE scores (n=25), but it did show significant improvement for program

participants. With 2020's much larger sample size (n=721), we can be certain of the significant increase in post-test scores for RBBB participants.

**Figure 4: Dibels DAZE Score Averages for 3rd Grade Participants; 2016 and 2020 data results shown**



\* Both examined years showed a statistically significant improvement from pre- to post-time measures. Paired samples t-tests were conducted with significance set at the  $\alpha = .05$  level.

### *Dibels ORF*

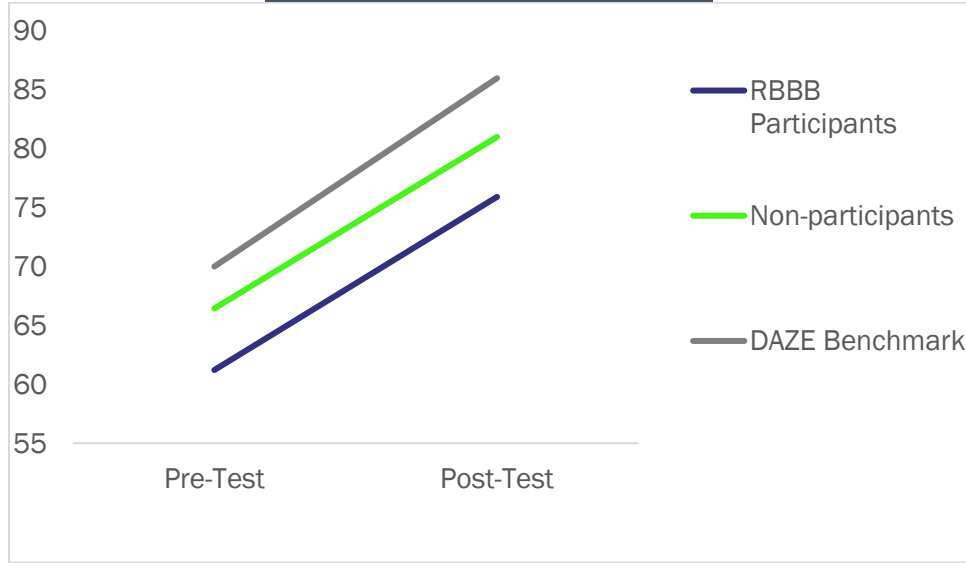
The DIBELS ORF (Oral Reading Fluency) was implemented throughout all semesters in this report's time period (Fall 2018 – Fall 2019) and was conducted in the following districts: Avondale, Buckeye, Littleton, Glendale, and Union. ORF data include both RBBB program participants and non-participants' scores. These scores are compared with the published ORF benchmark goal scores. In order to align with the benchmark goal scores, results are reported separately for all fall semesters and the spring semester. The same middle-of-year benchmark score was used twice to compare against the fall semesters' post-test and the spring semester's pre-test, as shown above in Table 3.

Analysis of the ORF scores also applied a matched-comparison group approach when analyzing the effect of RBBB program participation. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean ORF score of RBBB program participants, within one standard deviation of the participants' mean score.

### Fall semesters only:

For Readers in the RBBB program, during both Fall semesters (2018 and 2019), their DIBELS ORF scores increased almost 15 points, on average. The matched comparison group's score increased 14.5 points on average, while the benchmark goal shows a 16-point score increase. Even though the difference in increase between the participants and non-participants is small, it is statistically significant. The following visualization and table show these results.

**Figure 5: DIBELS ORF Average Scores, Pre and Post Test, Fall Semesters only**  
**(Chart axis truncated for detail.)**



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	61.21	75.92	14.71	n=392
Non-participants	66.44	81.01	14.57	n=689
ORF Benchmark Goal	70	86	16	

*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

RBBB participants started at pre-test time with a lower average ORF score than their matched comparison group, yet gained more than the comparison group's gains.

Again, the comparison groups in this report are imperfect. As well, the participants and the non-participant comparison group are similarly matched with respect to free and reduced lunch status. The RBBB participants are slightly more likely to be Hispanic/Latino than their non-participant counterparts.

#### Readers, Spring semester only:

During the Spring 2019 semester, 3<sup>rd</sup> grade Readers in the RBBB program increased their DIBELS ORF score over 16 points, while non-participants' score increase was a little over 13 points. This difference in score increase is statistically significant. The following visualization and table show these results.

***Figure 6: DIBELS ORF Average Scores, Pre and Post Test, Spring Semester only  
(Chart axis truncated for detail.)***



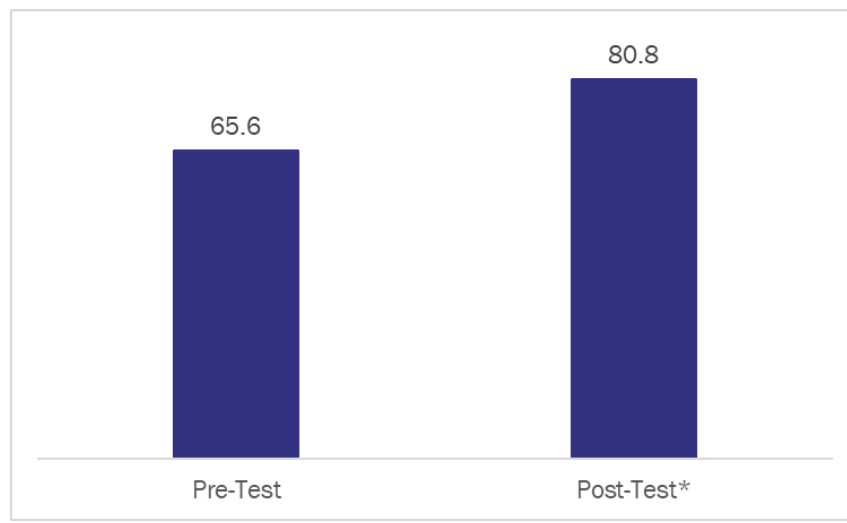
	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	76.16	92.54	16.38	n=161
Non-participants	81.39	94.57	13.18	n=341
ORF Benchmark Goal	86	100	14	

*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

The participants and the non-participant comparison group couldn't be compared with respect to free and reduced lunch status due to missing data, while the RBBB participant group has more African-American students than the comparison group during this semester. While neither participants nor non-participants meet the ORF benchmark goal scores, RBBB participants surpass the benchmark goal increase, gaining over 16 points.

Similar to the approach taken with the DAZE scores, another way of looking at the ORF data is to combine all the semesters and examine the RBBB participants' score increases over time. Figure 7 below shows results of 3<sup>rd</sup> grade participants' changes in average ORF scores. RBBB participating students demonstrate statistically significant improvement in their scores between pre-test time and post-test time.

*Figure 7: Dibels ORF Score Averages for 3rd Grade Participants (n=553)*



\* Current data show a statistically significant improvement from pre- to post-time measures. Paired samples t-tests were conducted with significance set at the  $\alpha = .05$  level.

### *Dibels Composite Score*

The DIBELS Composite Score was implemented throughout all semesters in this report's time period (Fall 2018 – Fall 2019) and was conducted in the following districts: Buckeye, Littleton, Glendale, and Union. Composite Score data include both RBBB program participants and non-participants' scores. These scores are compared with the published DIBELS Composite Score benchmark goal scores. In order to align with the benchmark goal scores, results are reported separately for all fall semesters and the spring semester. The same middle-of-year benchmark score was used twice to compare against the fall semesters' post-test and the spring semester's pre-test, as shown above in Table 3.

Analysis of the Composite scores also applied a matched-comparison group approach when analyzing the effect of RBBB program participation. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean ORF score of RBBB program participants, within one standard deviation of the participants' mean score.

Two repeated-measures general linear models ("GLM") were performed for two groups: Readers in the Spring, 2019 semester; and Readers in both of the Fall semesters (Fall 2018 or Fall 2019). Readers were separated into the Fall / Spring semester groups so as to compare their progress with Composite score benchmark goal scores. A GLM analysis tests whether the score increase for RBBB program participants is significantly greater than the score increase for comparison group students.

The data's sample size did not allow to test for school-level effects. Other statistical methods were explored (hierarchical linear modeling and propensity score matching) and found

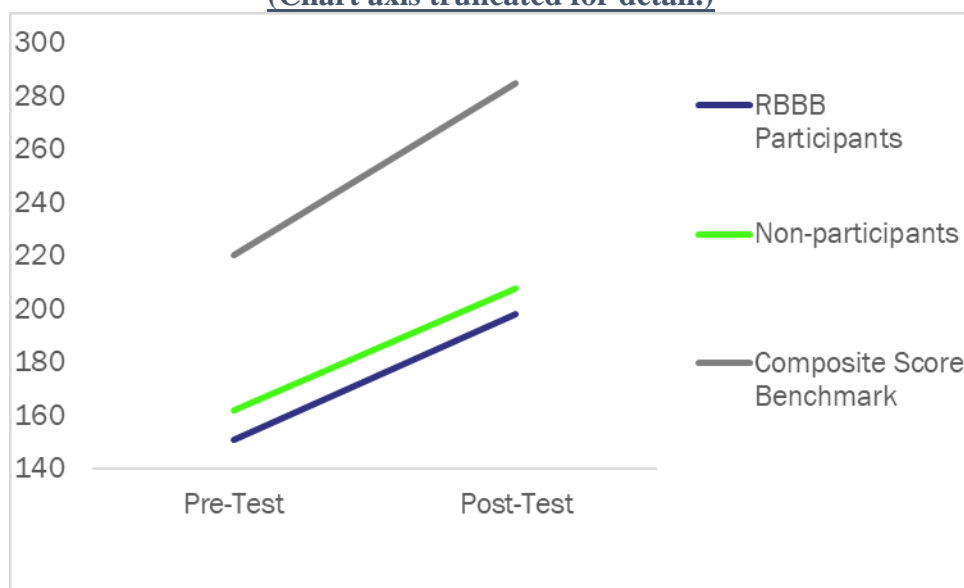


inappropriate for this data. A full description of the statistical testing methods follows in Appendix F.

#### Fall semesters only:

For Readers in the RBBB program, during both Fall semesters (2018 and 2019), their DIBELS Composite scores increased well over 47 points, on average. The matched comparison group's score increased almost 47 points on average, while the benchmark goal shows a 65-point score increase. The following visualization and table show these results.

**Figure 8: DIBELS Composite Score Averages, Pre and Post Test, Fall Semesters only**  
(Chart axis truncated for detail.)



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	151.19	199.06	47.87	n=215
Non-participants	162.51	209.3	46.79	n=329
Composite Score Benchmark Goal	220	285	65	

*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

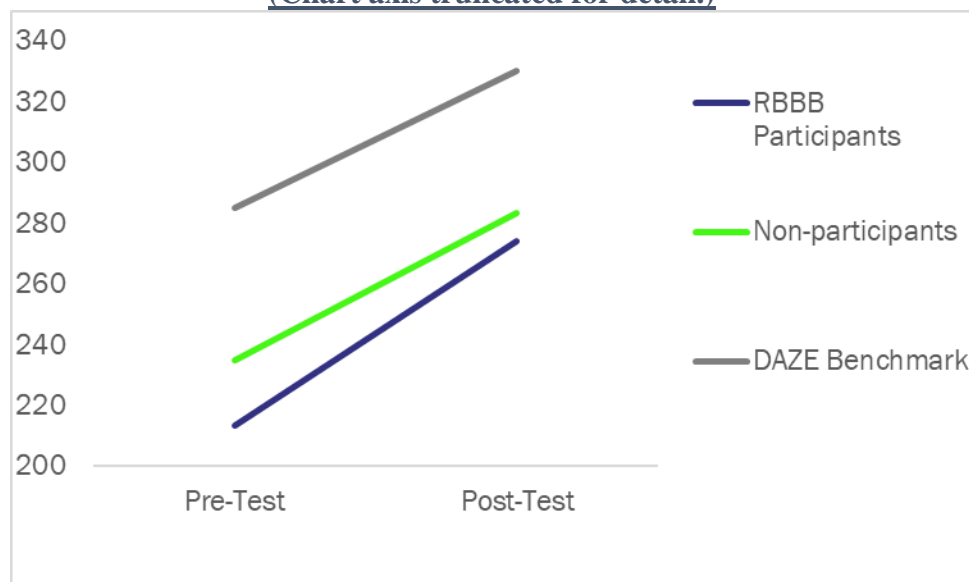
RBBB participants started at pre-test time with a lower average Composite score than their matched comparison group, and their score gains were higher than the comparison group's gains. These gains are statistically significantly greater than the non-participants' gains.

Again, the comparison groups in this report are imperfect. The RBBB participants are slightly more likely to be eligible for free lunch and slightly more likely to be Hispanic/Latino and African-American than their non-participant counterparts.

Readers, Spring semester only:

During the Spring 2019 semester, Readers in the RBBB program increased their DIBELS Composite scores over 60 points, while non-participants saw an increase of approximately 48 points. The following visualization and table show these results.

**Figure 9: DIBELS Composite Score Averages, Pre and Post Test, Spring Semester only**  
(Chart axis truncated for detail.)



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	213.47	273.87	60.4	n=174
Non-participants	235.02	283.3	48.28	n=305
Composite Score Benchmark Goal	285	330	45	

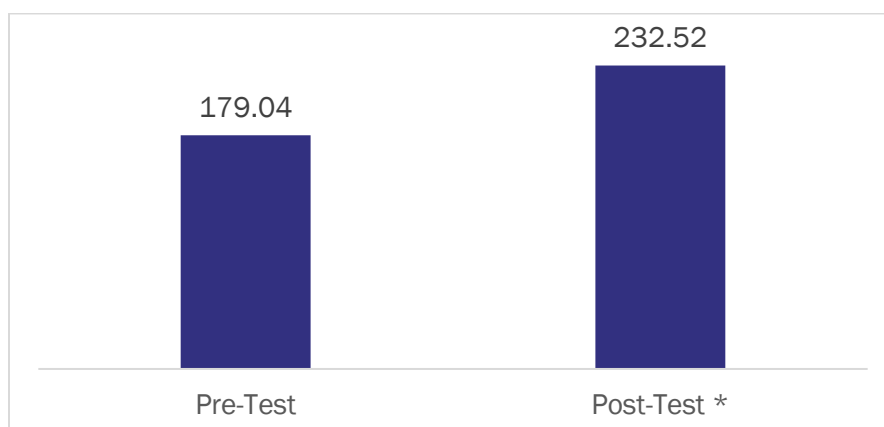
*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

Here, RBBB participants started at pre-test time with a lower Composite average score than their non-participant counterparts. RBBB participants' gains were much greater than their non-participant counterparts' gains, and this difference is statistically significant. Both groups increased their scores more than the recommended benchmark score increase of 45 points.

The participants and the non-participant comparison group couldn't be compared with respect to free and reduced lunch status due to missing data, while the RBBB participant group has slightly more African-American students than the comparison group during this semester.

Similar to the approach taken with the DAZE scores, another way of looking at the Composite Score data is to combine all the semesters and examine the RBBB participants' score increases over time. Figure 10 below shows results of 3<sup>rd</sup> grade participants' changes in average Composite scores. RBBB participating students demonstrate statistically significant improvement in their scores between pre-test time and post-test time.

Figure 10: Dibels Composite Score Averages for 3rd Grade Participants (n=777)



*\* Current data show a statistically significant improvement from pre- to post-time measures. Paired samples t-tests were conducted with significance set at the  $\alpha = .05$  level.*

### *Galileo*

The Galileo assessment was implemented throughout all semesters in this report's time period (Fall 2018 – Fall 2019) in the Avondale school district. Galileo score data include both RBBB program participants and non-participants' scores.

Analysis of the Galileo scores also applied a matched-comparison group approach when analyzing the effect of RBBB program participation. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean Galileo score of RBBB program participants, within one standard deviation of the participants' mean score.

Two repeated-measures general linear models ("GLMs") were performed for two groups: Readers in the Spring, 2019 semester; and Readers in both of the Fall semesters (Fall 2018 or Fall 2019). Readers were separated into the Fall / Spring semester groups so as to compare their progress with Galileo benchmark goal scores. A repeated-measures general linear models ("GLM") was performed, testing whether the score increase for RBBB program participants is significantly greater than the score increase for comparison group students.

### Fall semesters

For Readers in the RBBB program, their Galileo scores increased almost 42 points, on average. The matched comparison group's score increased by almost 20 points, and this difference is statistically significant. The 2016 data showed an 8-point increase in Galileo scores for RBBB program participants overall in all semesters.

As well, score benchmarks for the Galileo assessment exist at pre-test time (beginning of a Fall semester) and at mid-point time (end of a Fall semester and beginning of a Spring semester).

Score benchmarks do not exist for 3<sup>rd</sup> grade readers at post-test time, what would be the end of a Spring semester. The following visualization and table show these benchmarks and results.

**Figure 11: Galileo Score Averages, Pre and Post Test, Fall Semesters Only**  
(Chart axis truncated for detail.)

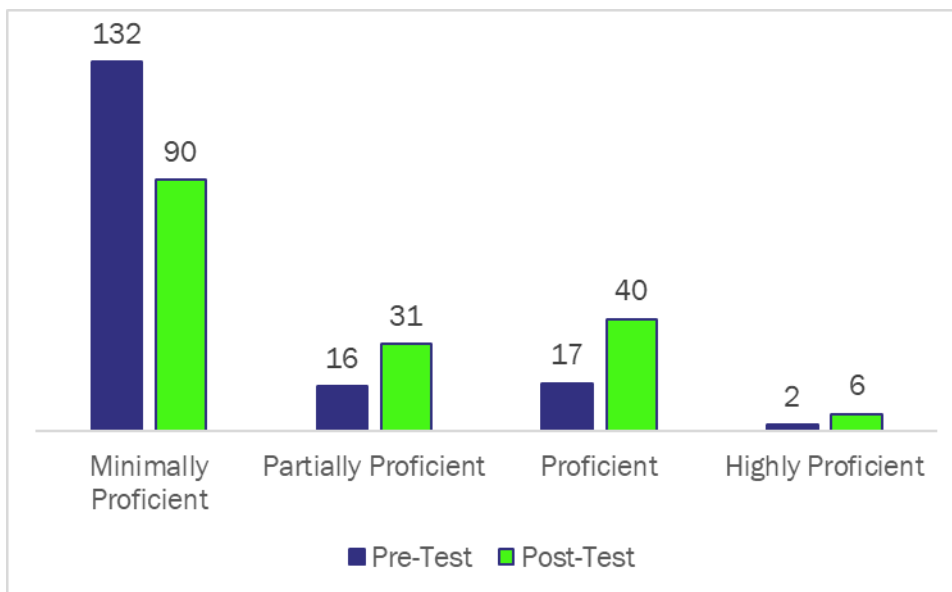


	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	686	728	42	n=167
Non-participants	690	710	20	n=654
"Proficient" Benchmark Goal Range	777-875	810-918		

*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

Another way of assessing the RBBB program participants' achievements for the Fall semesters is to examine how many students experienced movement from one benchmark score category to the next highest category. Galileo benchmark goals are provided as Minimally Proficient, Partially Proficient, Proficient, and Highly Proficient. The figure below shows the Fall participants' movement among these categories, in particular, that students increase their place into the higher proficiency categories as the semester occurs.

**Figure 12: Number of RBBB Students at Galileo Benchmark Goals during Fall Semesters**

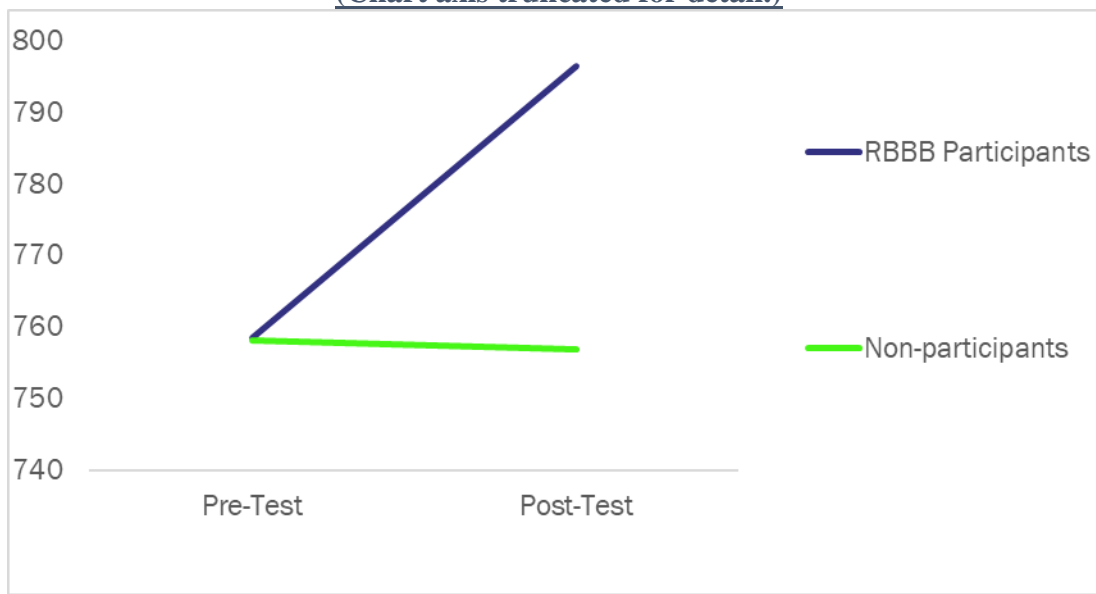


#### Spring semester only

For Readers in the RBBB program in the Spring 2019 semester, their Galileo scores increased approximately 38 points, on average. The matched comparison group's score decreased by one point. This difference approaches statistical significance with a p-value of .086. .

For the Spring semester, the only appropriate benchmark scores exist for the pre-test time. The following visualization and table show these benchmarks and results.

**Figure 13: Galileo Score Averages, Pre and Post Test, Spring Semester Only**  
(Chart axis truncated for detail.)



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	758.49	796.4	37.91	n=81
Non-participants	758.08	756.87	-1.21	n=182
“Proficient” Benchmark Goal Range	810-918	N/A		

*\* Participants’ score increase is nearly statistically significantly greater than the comparison group’s score increase at the  $\alpha=.05$  level, according to GLM test.*

The comparison groups for the Galileo assessment are imperfect, as the RBBB participants are slightly more likely to be eligible for free lunch and slightly more likely to be African-American and slightly less likely to be Hispanic/Latino than their non-participant counterparts.

### ORA

The ORA assessment was implemented throughout all semesters in this report’s time period (Fall 2018 – Fall 2019) in the Osborn school district. ORA score data include both RBBB program participants and non-participants’ scores.

Analysis of the ORA scores also applied a matched-comparison group approach when analyzing the effect of RBBB program participation. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean ORA score of RBBB program participants, within one standard deviation of the participants’ mean score.

A repeated-measures general linear models (“GLM”) was performed, testing whether the score increase for RBBB program participants is significantly greater than the score increase for comparison group students.

For Readers in the RBBB program, their ORA score increased by slightly over 3 points from pre-test to post-test time. This is statistically significantly greater than the matched comparison group’s score increase of approximately 2.5 points. Even though participants start out with lower scores on the ORA than their counterparts, their gains are greater. The following visualization and table show these results.

***Figure 14: ORA Average Scores, Pre and Post Test***  
***(Chart axis truncated for detail.)***



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	19.64	22.72	3.08	n=60
Non-participants	21.52	23.99	2.47	n=287

*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

The comparison group for the ORA data is not a perfect match. The RBBB participants are much more likely to be African-American and / or Hispanic/Latino than their non-participant counterparts. There was no data on students' free and reduced lunch status.

### *AIMSweb*

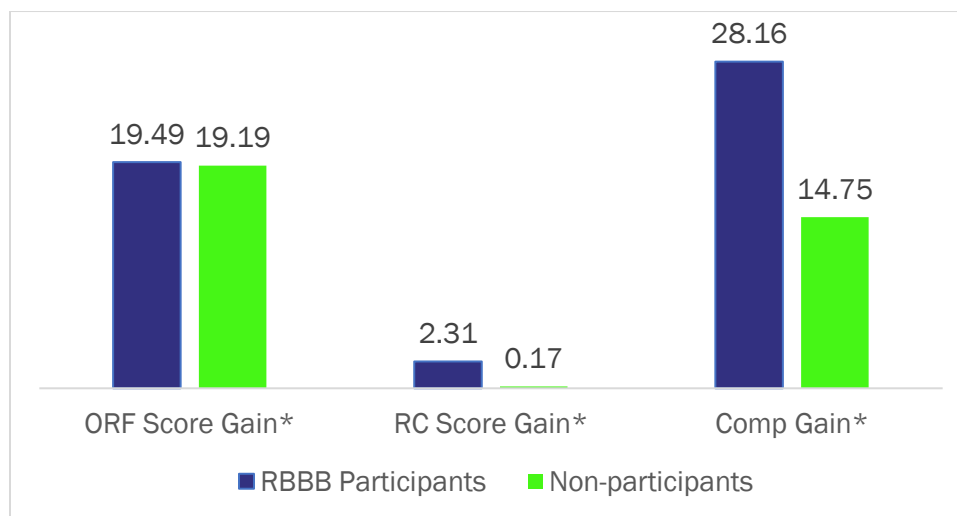
AIMSweb is administered in the Phoenix school district during all examined semesters. AIMSweb measures four primary areas: oral reading fluency, reading comprehension, vocabulary, and a composite score. There was not enough vocabulary assessment data (n=18) and so it is not included in this examination.

AIMSweb data contain RBBB participant and non-participant scores, and were examined with a matched, crafted comparison-group approach. GLMs were performed to test for statistical significance.

For RBBB participants, their increases in oral reading fluency, reading comprehension, and their composite score were all greater than the non-participants' gains in these areas. The RBBB

participants' gains are statistically significantly greater than their non-participant counterparts' gains. The following visualization shows these results.

**Figure 15: AIMSweb Score Gains from Pre-test to Post-test Time, Oral Reading Fluency, Reading Comprehension, and Composite Score**



*\* Participants' score increases are statistically significantly greater than the comparison group's score increases at the  $\alpha=.05$  level, according to GLM test.*

Most of the comparison group's ethnicity, racial, and lunch status data are missing and could not be compared with RBBB participants.

#### *NWEA MAP Test – RIT*

The Osborn school district implements the MAP test and reports RIT scores for all examined semesters.

RIT data include both RBBB program participants and non-participants' scores. These scores are compared with published RIT Reading Student Achievement Norms.<sup>14</sup> In order to align with the benchmark goal scores, results are reported separately for all fall semesters and the spring semester. The same middle-of-year benchmark score was used twice to compare against the fall semesters' post-test and the spring semester's pre-test. The following table describes this alignment:

<sup>14</sup> See 2020 NWEA MAP Growth Normative Data Overview:  
<https://teach.mapnwea.org/impl/MAPGrowthNormativeDataOverview.pdf>



**Table 6: Alignment of RBBB Data with RIT Norms**

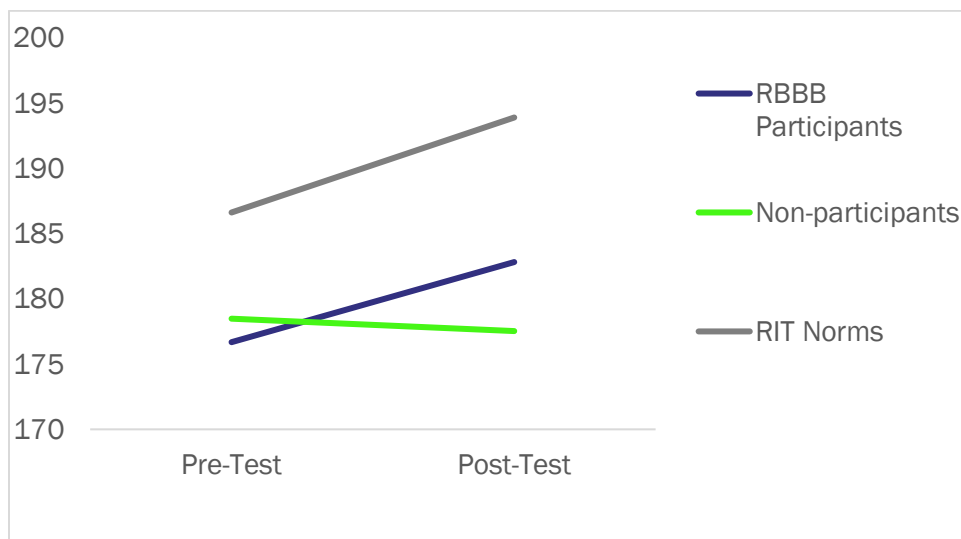
<b>RBBB Data Time</b>	<b>DIBELS ORF Benchmark Time</b>
Fall Semester, Pre-Test	Beginning of Year
Fall Semester, Post-Test	Middle of Year
Spring Semester, Pre-Test	Middle of Year
Spring Semester, Post-Test	End of Year

This analysis applies a matched-comparison group approach when analyzing the effect of RBBB program participation on RIT scores. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean RIT score of RBBB program participants, within one standard deviation of the participants' mean score.

**Fall semesters only:**

For Readers in the RBBB program, during both Fall semesters (2018 and 2019), their RIT scores increased over 6 points, on average. The matched comparison group's score showed a slight decrease, while the benchmark goal shows an increase of over 7 points. The participants' gain in RIT scores are not statistically significantly different compared to the non-participants' gains, likely due to the small sample size of participants (n=44). The following visualization and table show these results.

**Figure 16: RIT Average Scores, Pre and Post Test, Fall Semesters only**



	<b>Pre Test</b>	<b>Post Test</b>	<b>Score Increase</b>	<b>Group Size</b>
RBBB Participants	176.68	182.82	6.14	n=44
Non-participants	178.48	177.53	-0.95	n=219
RIT Norms	186.62	193.9	7.28	

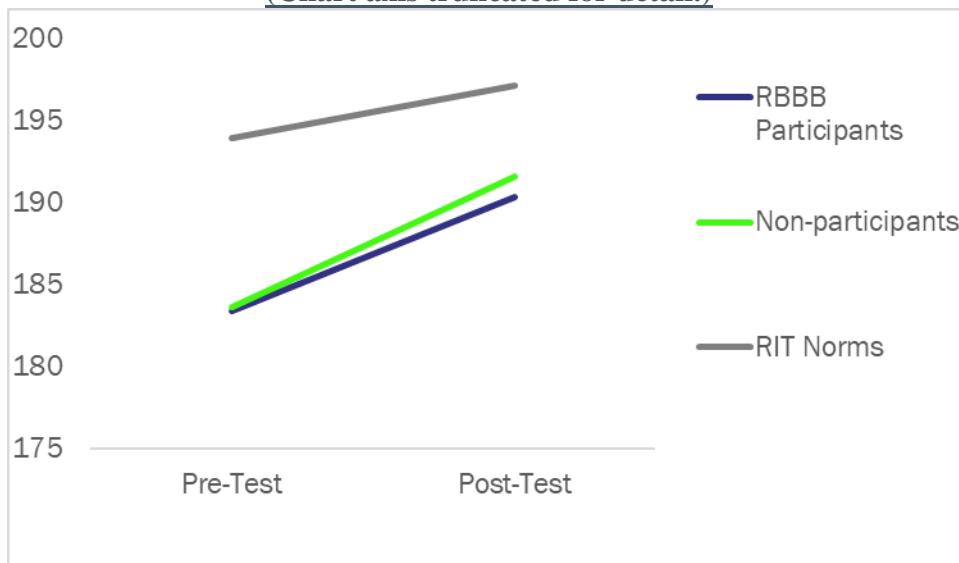
RBBB participants started at pre-test time with a slightly lower average RIT score than their matched comparison group, but their gains were much greater. Yet, neither group gained nor met the RIT benchmark score norms.

It is important to mention that the comparison group of non-participants is an imperfect comparison group. It is much larger than the participant group, and sample sizes can impact results when comparing averages, as shown in the table above. RBBB participants are more likely to be African-American than the non-participants. No data were available on free lunch status.

#### Spring semester only:

For Readers in the RBBB program, during the Spring 2019 semester, their RIT scores increased almost 7 points, on average. The matched comparison group's score increased well over 7 points, while the benchmark goal shows an increase of just over 3 points. The following visualization and table show these results.

***Figure 17: RIT Average Scores, Pre and Post Test, Spring Semester only***  
(Chart axis truncated for detail.)



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	183.37	190.33	6.96	n=27
Non-participants	183.63	191.58	7.95	n=72
RIT Norms	193.9	197.12	3.22	

RBBB participants started at pre-test time with almost the same average RIT score than their matched comparison group and did not gain as much as their comparison group counterparts

during this spring semester. The RBBB participants during this spring semester were more likely to be African-American than the non-participants.

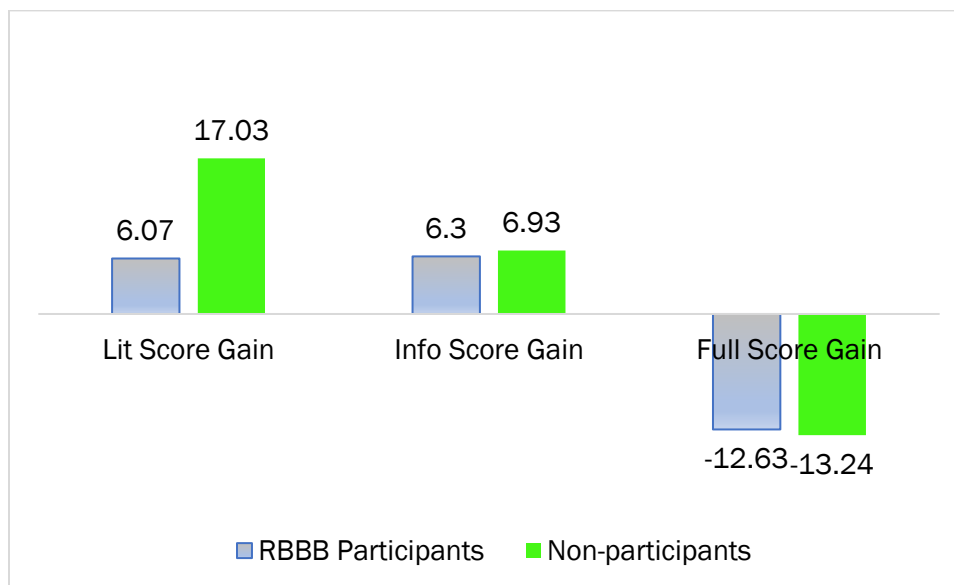
### *iReady*

iReady is administered in the Alhambra school district only in the Spring, 2019 and Fall, 2019 semesters. iReady is measured in three areas: Literature, Informational, and Full-Overall in the Spring semester, and Full-Overall in the Fall, 2019 semester.

iReady data contain RBBB participant and non-participant scores, and were examined with a matched, crafted comparison-group approach. GLMs were performed to test for statistical significance.

For RBBB participants, their increases in the Literature assessment score are less than the non-participants' increases, and their gains in the Informational score are almost equal to the non-participants' gains. For both participants and non-participants, their Full-Overall score decreased between pre-test and post-test time. The following visualization shows these results.

**Figure 18: iReady Score Gains from Pre-test to Post-test Time, Lit, Info, and Full**



RBBB participants and the non-participant comparison group appear to be similarly matched on ethnicity, racial, and free lunch status.

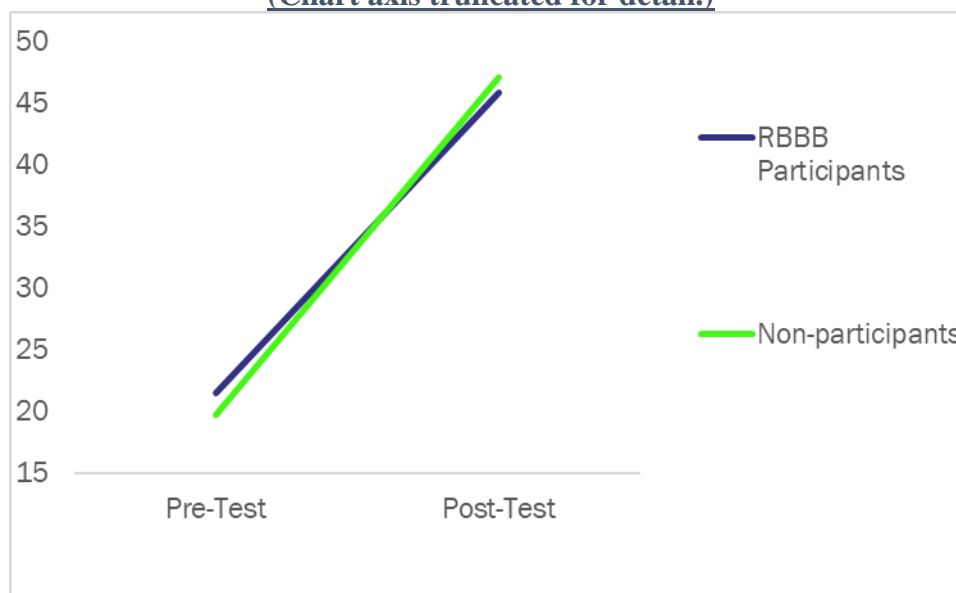
### *Illuminate*

The Illuminate assessment is administered in the Tolleson and Littleton school districts during all of the examined semesters. Illuminate data contain RBBB participant and non-participant scores,

and were examined with a matched, crafted comparison-group approach. GLMs were performed to test for statistical significance.

For Readers in the RBBB program, their Illuminate scores increased over 24 points, on average. The matched comparison group's score increased over 27 points, however. The following visualization and table show these results.

***Figure 19: Illuminate Average Scores, Pre and Post Test***  
(Chart axis truncated for detail.)



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	21.45	45.90	24.45	n=166
Non-participants	19.68	47.15	27.47	n=643

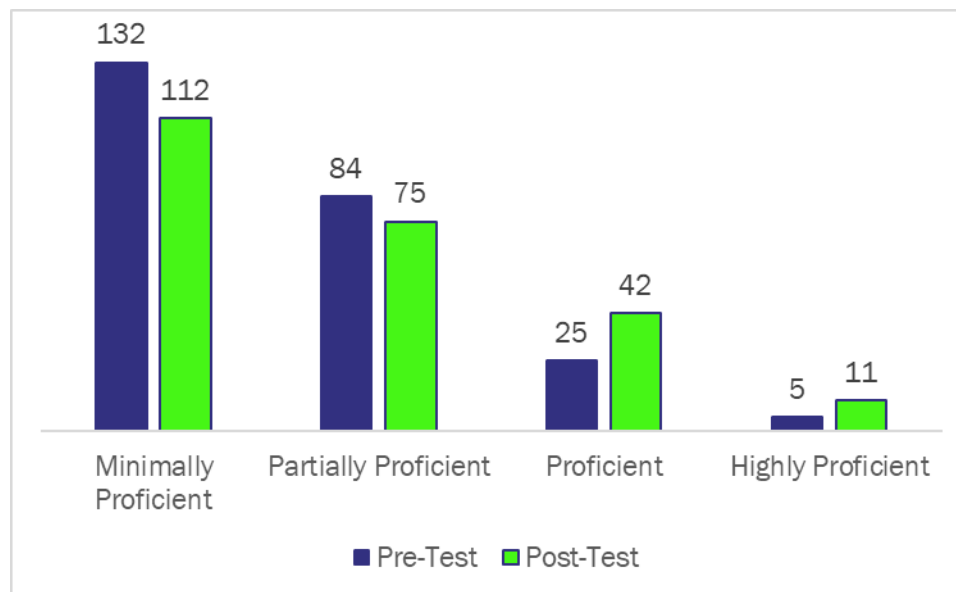
RBBB participants and the non-participant comparison group appear to be similarly matched on ethnicity, racial, and free lunch status.

#### **FAME**

The FAME assessment is administered in the Avondale school district during all of the semesters examined in this report. The data contain RBBB participant and non-participant scores. FAME is measured qualitatively as Minimally Proficient, Partially Proficient, Proficient, and Highly Proficient. These qualitative measures can be compared, as shown below.

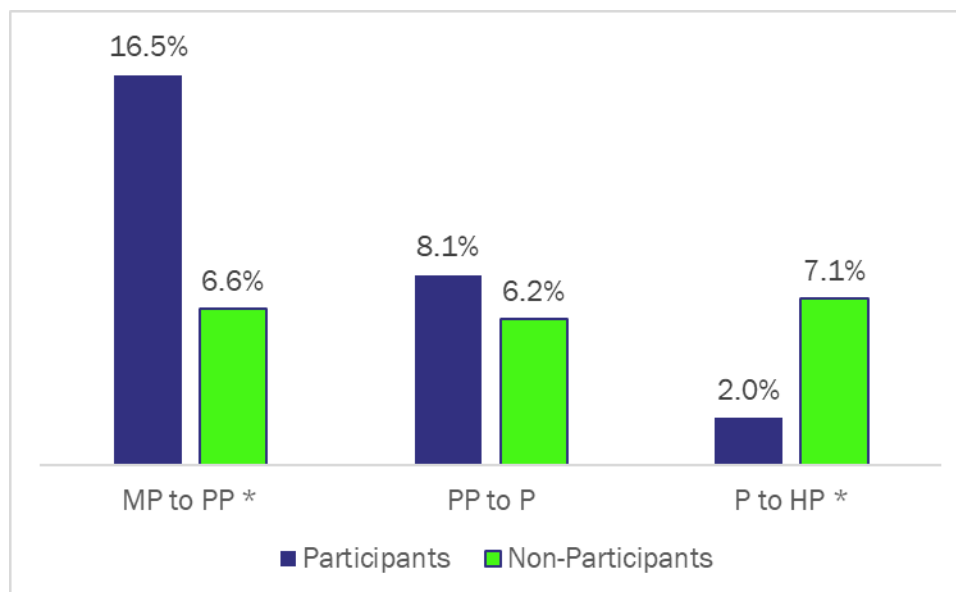
For Readers in the RBBB program, as they progress throughout their semesters, they are more likely to move into the “Proficient” category or the “Highly Proficient” category, as shown in the figure below.

***Figure 20: Number of RBBB Participants in FAME Category at Pre and Post Test Time***



It is also possible to examine the proportion of participants who moved to the next FAME category between pre and post-test time and compare this with the same proportion of non-participants. The figure below shows that 16.5% of RBBB program participants moved from “Minimally Proficient” to “Partially Proficient” while only 6.6% of non-participants did. This difference is statistically significant. There was no significant difference, however, between participants and non-participants moving from “Partially Proficient” to “Proficient.” And, regarding the move from “Proficient” to “Highly Proficient,” non-participants made a greater jump than participants did.

**Figure 21: Percentage of RBBB Participants and Non-Participants Moving FAME Categories from Pre Test to Post Test Time**

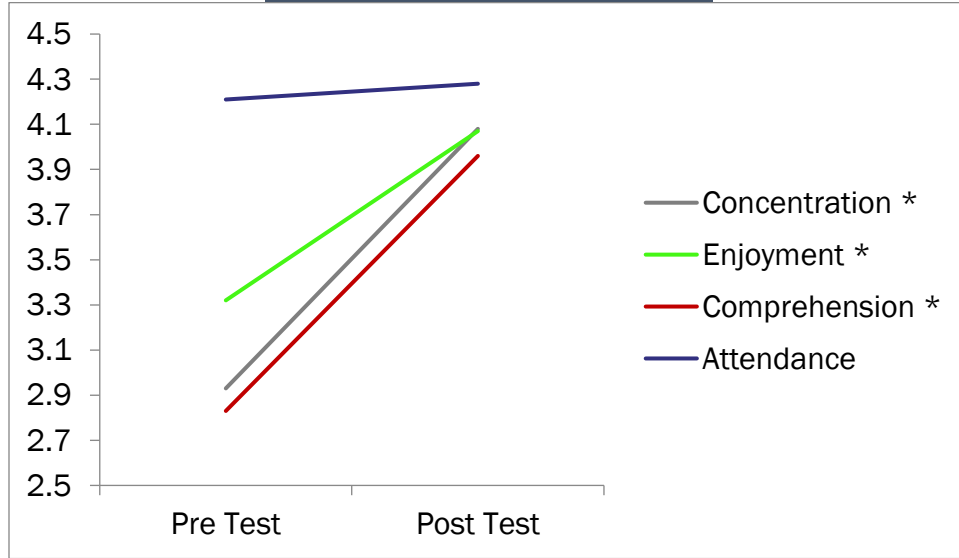


*\* Denotes a statistically significant difference in proportions between RBBB participants and non-participants, at the  $\alpha = .05$  level, according to Z-scores tests of proportions.*

#### *Teacher Evaluation of Students – “Reader Teacher Survey”*

RBBB collects teacher perceptions of students from the following school districts: Alhambra, Avondale, Buckeye, Littleton, Osborn, Phoenix, Tolleson, and Union. Teachers evaluated their 3<sup>rd</sup> grade students who participated in the RBBB program in four areas – Focus, Enjoyment of Reading, Reading Comprehension, and Attendance. Figure 21 below shows the average score for participants from the beginning of the semester (“pre”) to the end of the semester (“post”). On average, participants improved on all four subscales to a statistically significant extent. The subscales were scored from 1 to 5, with a higher score denoting greater teacher agreement with the student’s achievement. The following visualization and table show these results.

***Figure 22: Teacher Evaluations of 3rd Grade Participants***  
***(Chart axis truncated for detail.)***



	Pre Test	Post Test	Score Increase
Concentration	2.93	4.08	1.15 *
Enjoyment	3.32	4.07	0.75 *
Comprehension	2.83	3.96	1.13 *
Attendance	4.21	4.28	0.07

*\* Denotes a statistically significant improvement from pre- to post-time measures, at the  $\alpha = .05$  level, according to paired-samples t-test.*

Beyond the teachers' perception of attendance, the data allow for testing correlation between attendance and increases in concentration, enjoyment, and comprehension. On average, attendance was positively correlated with greater increases on the teacher evaluation scores. The more sessions a student attended, the higher their increase in teacher evaluation score from pre-test to post-test time. Attendance was significantly positively correlated<sup>15</sup> with an increase in enjoyment of reading and positively correlated with increased concentration and comprehension.

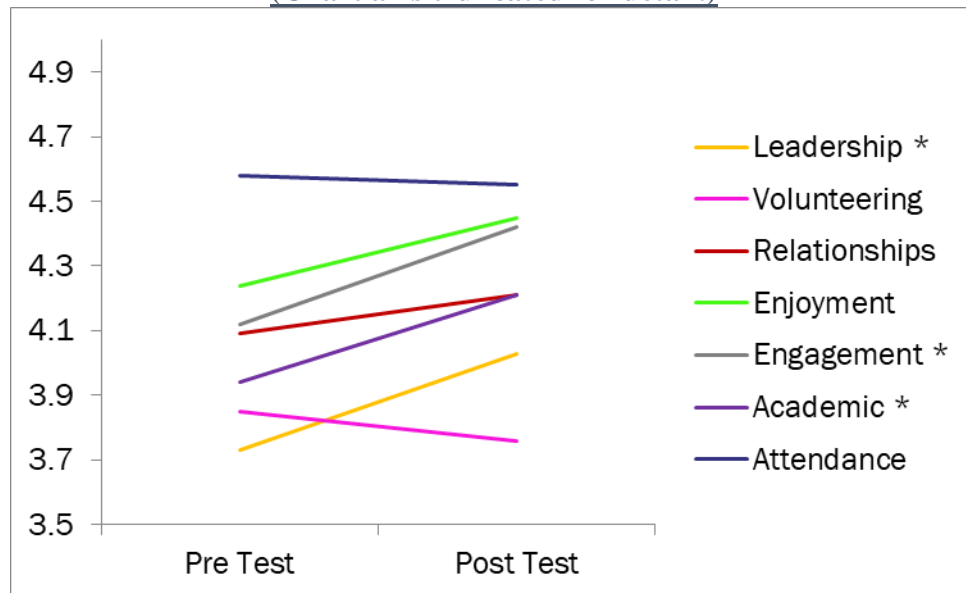
#### *Teacher Evaluation of Leaders – “Leader Teacher Survey”*

RBBB collects teacher perceptions of students from the following school districts: Alhambra, Avondale, Buckeye, Littleton, Osborn, and Phoenix. Teachers evaluated the 8<sup>th</sup> grade Leaders who participated in the RBBB program in seven areas – Leadership skills, Volunteer initiative, Positive relationships, Enjoyment of school, Engagement in the classroom, Academic expectations, and Attendance. Figure 22 below shows the average score for participants from the

<sup>15</sup> Bivariate Pearson correlation found to be statistically significant at the  $\alpha = .05$  level.

beginning of the semester (“pre”) to the end of the semester (“post”). Participants improved on 5 of the 7 subscales, and statistically significantly improved on the subscales measuring their Leadership, Engagement, and Academic achievement. The subscales were scored from 1 to 5, with a higher score denoting greater teacher agreement with the student’s achievement. The following visualization and table show these results.

***Figure 23: Teacher Evaluations of 8th Grade Participants***  
(Chart axis truncated for detail.)



	Pre Test	Post Test	Score Increase
Leadership	3.73	4.03	0.3 *
Volunteering	3.85	3.76	-0.09
Relationships	4.09	4.21	0.12
Enjoyment	4.24	4.45	0.21
Engagement	4.12	4.42	0.30 *
Academic	3.94	4.21	0.27 *
Attendance	4.58	4.55	-0.03

*\* Denotes a statistically significant improvement from pre- to post-time measures, at the  $\alpha = .05$  level, according to paired-samples t-test.*

For the RBBB Leaders’ teacher evaluation data, a completed pre and post-test must be in the data for each student in order to examine their improvement. This was the case in the data for 33 students.

Beyond the teachers’ perception of attendance, the data allow for testing for correlations between attendance and increases in leadership, volunteering, relationships, enjoyment, engagement, and academic attainment. On average, attendance was positively correlated with greater increases in all of these measures on the teacher evaluation scores except for Relationships, however, these correlations were not statistically significant.



### *Reading Leader Survey*

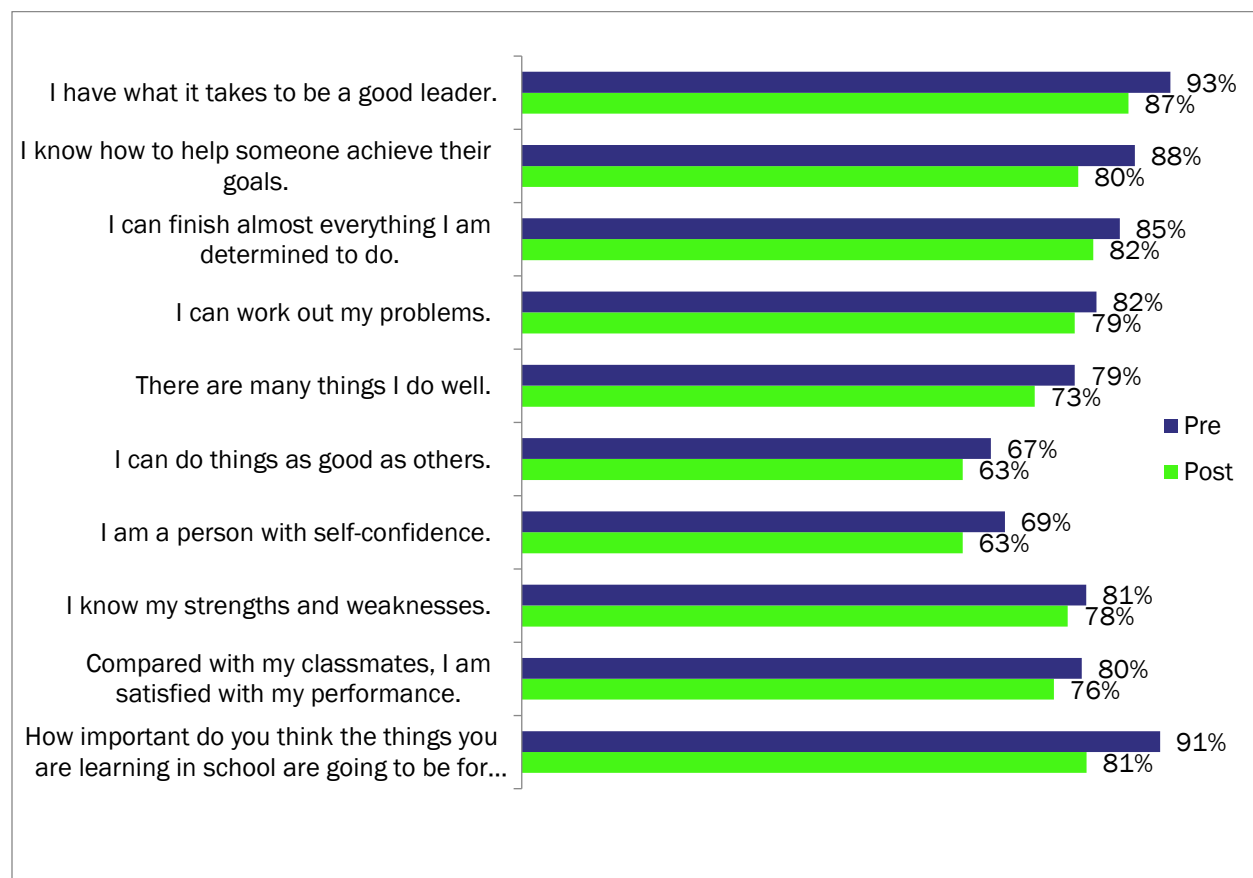
Not all schools have pretest data, and not all schools have post-test data. Like the teacher evaluation data, a completed pre-test and post-test must be present in order to assess student improvement. This was the case in the data for 235 students.

For these students, their average post-test time score was similar, and slightly decreased, from their average pre-test time score. Questions on the Reading Leader Survey are scaled from 1 to 5, with 5 indicating greater agreement with the positive statements. Overall, students went from an average score of **4.2** to **4.1** throughout their time in the program.

The data also allow a specific look at each question to see whether there are large variations in student response from pre-test to post-test time. After examining all of the questions, all of the questions on the survey showed either **no change or a very slight decrease** in student responses. None of the questions showed improvement from pre-test to post-test time.

Figure 24 below reports the percentage of students responding that they either “Strongly Agree” or “Agree” with each statement. Overall, students didn’t consistently or significantly change in their agreement with the statements from pre-test to post-test time.

***Figure 24: Percentage of Leaders either Strongly Agreeing or Agreeing with each Statement, RLS***



After examining score differences by attendance, there does exist a **positive correlation** between program attendance and an average score increase from pre-test to post-test time.

~~

Overall, strong, significant evidence exists of RBBB program effectiveness on 3<sup>rd</sup> grade Readers. The following figure summarizes the impact shown among the different assessments.

***Figure 25: Summary of Data Results for 3<sup>rd</sup> Grade Readers, by Assessment Instrument***

	Little's	Bigs
RSPS	Strong Effect	Not Tested / N/A
DIBELS DAZE	Moderate Effect	Not Tested / N/A
DIBELS ORF	Strong Effect	Not Tested / N/A
DIBELS Composite	Strong Effect	Not Tested / N/A
Galileo	Strong Effect	Not Tested / N/A
ORA	Strong Effect	Not Tested / N/A
AIMSweb	Strong Effect	Not Tested / N/A
NWEA MAP - RIT	Mixed Effect	Not Tested / N/A
iReady	No Effect	Not Tested / N/A
Illuminate	Moderate Effect	Not Tested / N/A
FAME	Mixed Effect	Not Tested / N/A
Teacher Eval of Students	Strong Effect	Moderate Effect
RLS	Not Tested / N/A	No Effect

Key:

Strong Effect
Moderate Effect
Mixed Effect
No Effect
Not Tested / N/A

## Results – 2<sup>nd</sup> Grade Pilot Program

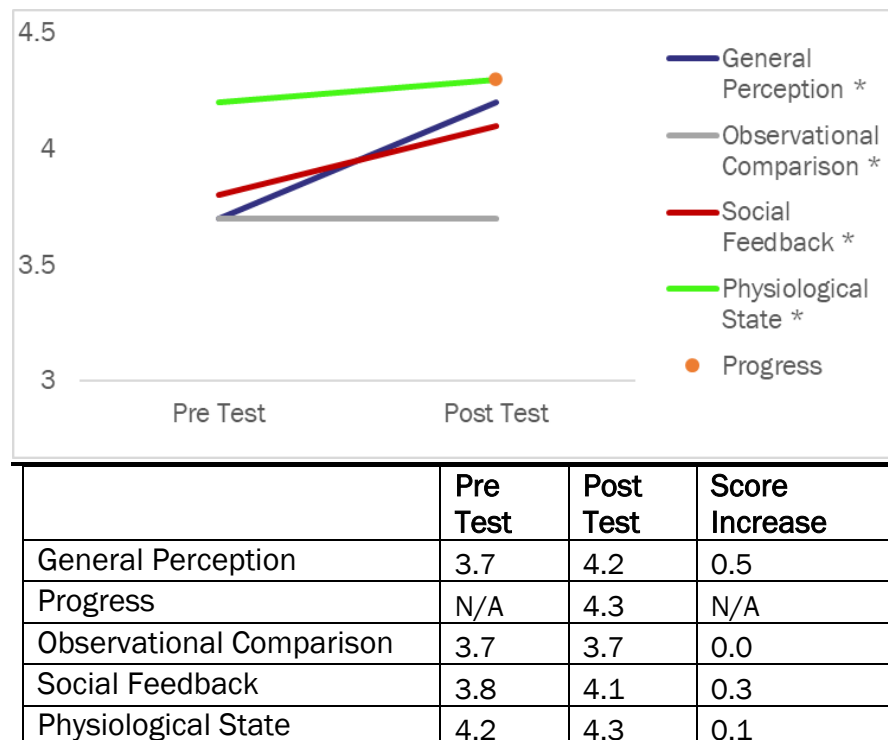
RBBB piloted their program with 2<sup>nd</sup>-grade students during the semesters covered in this report in the Phoenix, Avondale, and Glendale school districts. This section details the results of that pilot.

### *Reader Self-Perception Scale (“RSPS”)*

For 2<sup>nd</sup> grade Readers, the RSPS was implemented in following districts: Avondale, Glendale, and Phoenix.

Figure xx below shows results of 2<sup>nd</sup> grade participants’ changes in scores on the RSPS. Second graders did not show statistically significant changes in their RSPS scores from pre-test to post-test time. A small sample size (n=38) likely explains this.

**Figure 26: RSPS Subscale Component Score Averages for 3rd Grade Participants (Scored from 1 – 5; Chart axis truncated for detail.)**



\* Paired samples t-tests were conducted with significance set at the  $\alpha = .05$  level.

This analysis also explored the potential impact that attendance in RBBB sessions could have on progress. For the RSPS, on average, the number of RBBB program sessions attended did not notably correlate with greater increases on the RSPS subscale scores.<sup>16</sup>

<sup>16</sup> Bivariate Pearson correlation with statistical significance measured at the  $\alpha = .05$  level.

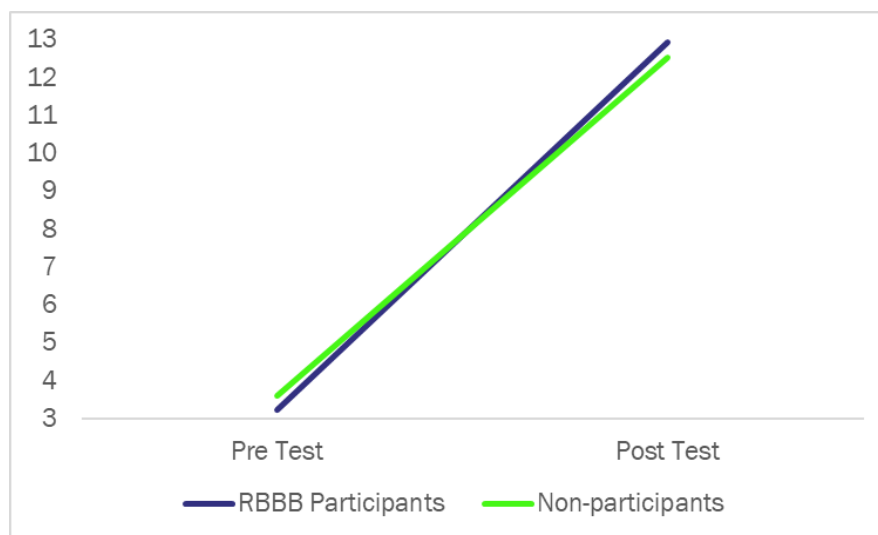
## DIBELS DAZE

The DIBELS DAZE was used to assess achievement for the 2<sup>nd</sup> grade pilot program during the Fall, 2019 semester in the Avondale school districts. DAZE data include both RBBB program participants and non-participants' scores. For 3<sup>rd</sup>-grade Readers, these scores could be compared with published DAZE benchmark goal scores.<sup>17</sup> However, benchmark scores are not published for 2<sup>nd</sup>-grade students.

This analysis applies a matched-comparison group approach when analyzing the effect of RBBB program participation on DAZE scores. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean DAZE score of RBBB program participants, within one standard deviation of the participants' mean score.<sup>18</sup>

For Readers in the RBBB program, their DIBELS DAZE scores increased 9.7 points, on average. The matched comparison group's score increased less than 9 points on average, and this difference is statistically significant. The following visualization and table show these results.

**Figure 27: DIBELS DAZE Average Scores, Pre and Post Test, Fall Semesters only**



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	3.23	12.92	9.69	n=13
Non-participants	3.58	12.53	8.95	n=38

*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

<sup>17</sup> All Dibels benchmark scores can be found here:

<https://www.nncsk12.org/site/handlers/filedownload.ashx?moduleinstanceid=6044&dataid=5471&FileName=DIBELSNextBenchmarkGoals-7.pdf>

<sup>18</sup> A matched comparison-group approach follows the lead of this white paper: "Matched Comparison Group Design: An Evaluation Brief for Educational Stakeholders." Jan. 2017, Hanita et al.

[https://www.edc.org/sites/default/files/uploads/matched\\_comparison\\_group\\_design.pdf](https://www.edc.org/sites/default/files/uploads/matched_comparison_group_design.pdf)

While these results reveal a statistically significant difference, the sample size of participants is very small.

### *Dibels ORF*

The DIBELS ORF (Oral Reading Fluency) was implemented throughout all semesters in this report's time period (Fall 2018 – Fall 2019) and was conducted in the Avondale and Glendale school districts. ORF data include both RBBB program participants and non-participants' scores. These scores are compared with the published ORF benchmark goal scores.

However, there are only 5 participants in the 2<sup>nd</sup> grade pilot during a Spring semester. Therefore, results are reported for Fall semester students only, and compared to the ORF Fall benchmarks.

Analysis of the ORF scores also applied a matched-comparison group approach when analyzing the effect of RBBB program participation. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean ORF score of RBBB program participants, within one standard deviation of the participants' mean score. GLM analyses were performed to test for statistical significance.

#### Fall semesters only:

For 2<sup>nd</sup> grade Readers in the RBBB program, during both Fall semesters (2018 and 2019), their DIBELS ORF scores increased over 28 points, on average. The matched comparison group's score increased 24.5 points on average, while the benchmark goal shows a 20-point score increase. This difference in increase between the participants and non-participants is statistically significant. The following visualization and table show these results.

**Figure 28: DIBELS ORF Average Scores, Pre and Post Test, Fall Semesters only**  
**(Chart axis truncated for detail.)**



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	44.19	72.81	28.62	n=36
Non-participants	46.43	70.89	24.46	n=142
ORF Benchmark Goal	52	72	20	

*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

RBBB participants started at pre-test time with a lower average ORF score than their matched comparison group, yet gained more than the comparison group's gains and surpassed the benchmark goal gain.

#### *Dibels Composite Score*

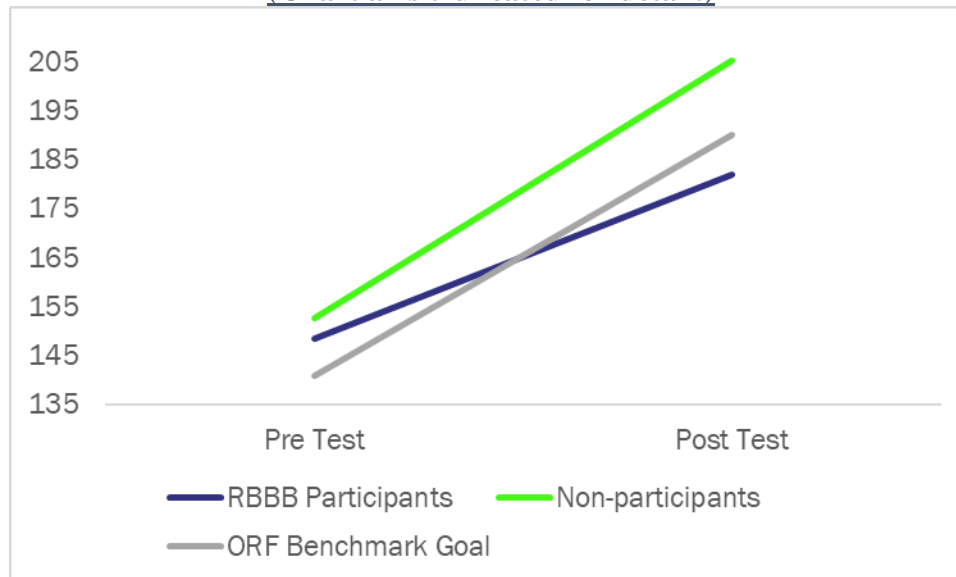
The DIBELS Composite Score was implemented throughout the Fall, 2019 semester in the Glendale school district for 2<sup>nd</sup>-grade participants. Composite Score data include both RBBB program participants and non-participants' scores. These scores are compared with the published DIBELS Composite Score benchmark goal scores for a fall semester.

Analysis of the Composite scores also applied a matched-comparison group approach when analyzing the effect of RBBB program participation. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean ORF score of RBBB program participants, within one standard deviation of the participants' mean score. GLM analyses were performed to test for statistical significance.

#### Fall semesters only:

For 2<sup>nd</sup> grade Readers in the RBBB program, during the Fall 2019 semester, their DIBELS Composite scores increased over 33 points, on average. However, the matched comparison group's score increased almost 53 points on average, while the benchmark goal shows a 49-point score increase. The following visualization and table show these results.

***Figure 29: DIBELS Composite Score Averages, Pre and Post Test, Fall Semesters only  
(Chart axis truncated for detail.)***



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	148.38	181.85	33.47	n=13
Non-participants	152.57	205.14	52.57	n=14
Composite Score Benchmark Goal	141	190	49	

The comparison group in this example is likely not a perfect match, and the sample sizes of both participants and non-participants is incredibly small.

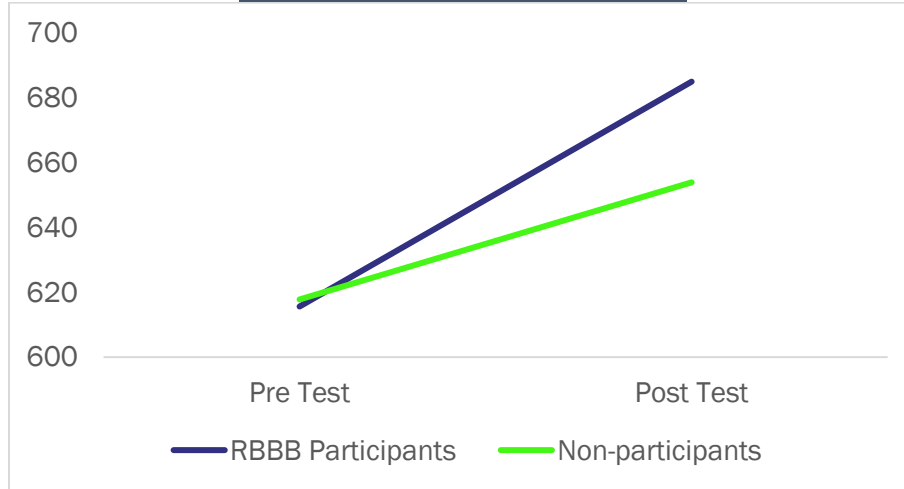
### *Galileo*

The Galileo assessment was implemented in the Fall, 2018 and Spring, 2019 semesters in the Avondale school district. Galileo score data include both RBBB program participants and non-participants' scores. Benchmark goal ranges are not available for 2<sup>nd</sup> grade students for Galileo.

Analysis of the Galileo scores also applied a matched-comparison group approach when analyzing the effect of RBBB program participation. A comparison group of non-program participants was crafted from available data to match the pre-test starting point mean Galileo score of RBBB program participants, within one standard deviation of the participants' mean score. GLM analyses were performed to test for statistical significance.

For 2<sup>nd</sup> grade Readers in the RBBB program, their Galileo scores increased over 69 points, on average. The matched comparison group's score increased by 36 points, and this difference is statistically significant, in spite of a very small sample size. The following visualization and table show these benchmarks and results.

***Figure 30: Galileo Score Averages, Pre and Post Test, Fall Semesters Only  
(Chart axis truncated for detail.)***



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	615.6	684.87	69.27	n=15
Non-participants	617.77	653.86	36.09	n=35

*\* Participants' score increase is statistically significantly greater than the comparison group's score increase at the  $\alpha=.05$  level, according to GLM test.*

### *AIMSweb*

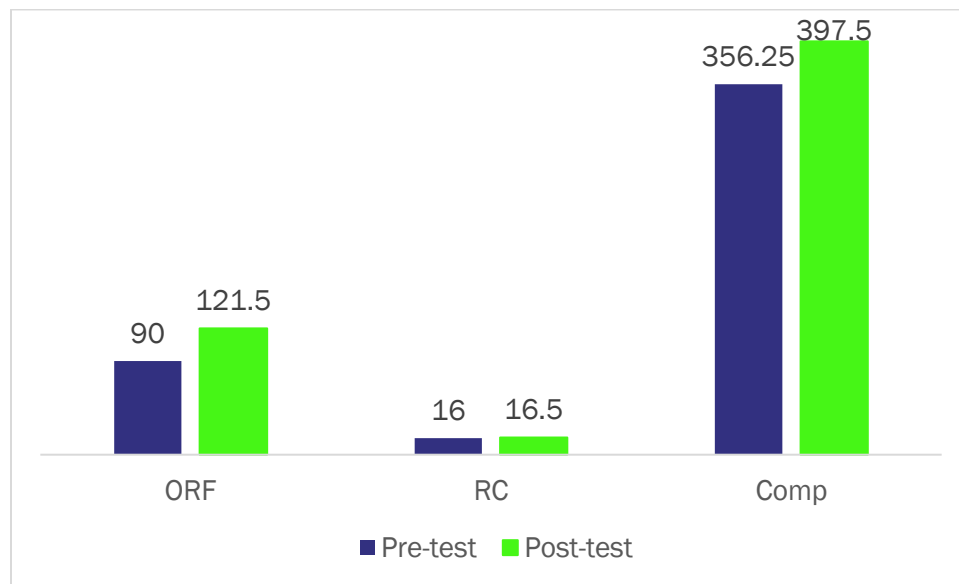
In the Fall of 2019, 4 Phoenix district 2<sup>nd</sup>-grade student Readers took the AIMSweb assessment.

AIMSweb is administered in the Phoenix school district during all examined semesters. AIMSweb measures four primary areas: oral reading fluency, reading comprehension, vocabulary, and a composite score. There was no vocabulary assessment data.

Given the tiny sample size for this assessment, the average scores are visualized in the figures below.



***Figure 31: AIMSweb Scores from Pre-test to Post-test Time, 2<sup>nd</sup> Grade, n=4***

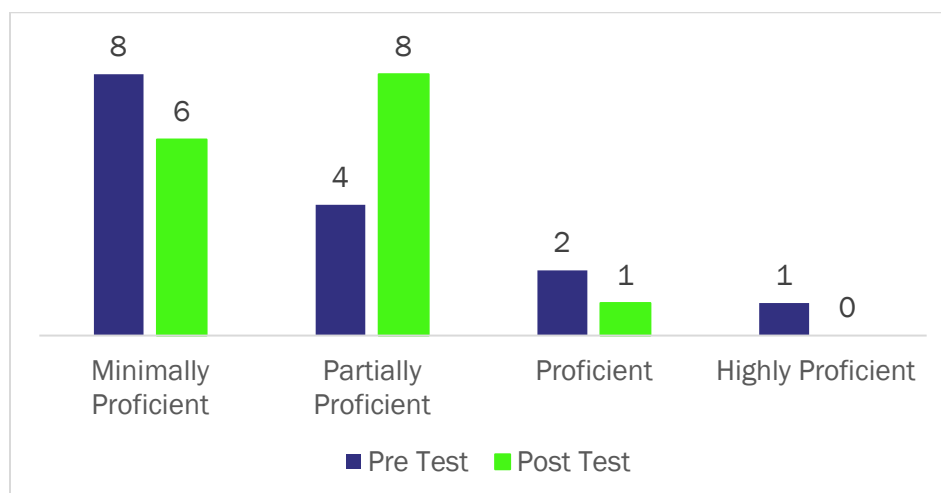


#### ***FAME***

The FAME assessment was administered in the Avondale school district to 2<sup>nd</sup>-grade Readers during the 2018-19 school year. The data contain RBBB participant and non-participant scores. FAME is measured qualitatively as Minimally Proficient, Partially Proficient, Proficient, and Highly Proficient. These qualitative measures can be compared, as shown below.

Only 15 2<sup>nd</sup> grade Readers were part of the RBBB program and took the FAME assessment. For Readers in the RBBB program, as they progress throughout their semesters, they are more likely to move into the “Partially Proficient” category, as shown in the figure below.

***Figure 32: Number of RBBB 2<sup>nd</sup> Grade Participants in FAME Category at Pre and Post Test Time (n=15)***



Due to the small sample size of participants, further examination with non-participants doesn't provide valid information about the gains made by the 2<sup>nd</sup>-grade participants.

#### *Teacher Evaluation of Students – “Reader Teacher Survey”*

RBBB collects teacher perceptions of 2<sup>nd</sup> grade Reader students from the Avondale school district throughout all semesters in this report. Teachers evaluated their 3<sup>rd</sup> grade students who participated in the RBBB program in four areas – Focus, Enjoyment of Reading, Reading Comprehension, and Attendance.

However, the data only contain 3 students who received both pre-test and post-test assessments. Many other 2<sup>nd</sup> grade pilot students received one or the other – either a pre-test assessment or a post-test assessment. The figure below shows the average score for participants from the beginning of the semester (“pre”) to the end of the semester (“post”), but these score values are of unpaired assessments. On average, participants improved on three of the four subscales. The subscales were scored from 1 to 5, with a higher score denoting greater teacher agreement with the student's achievement.

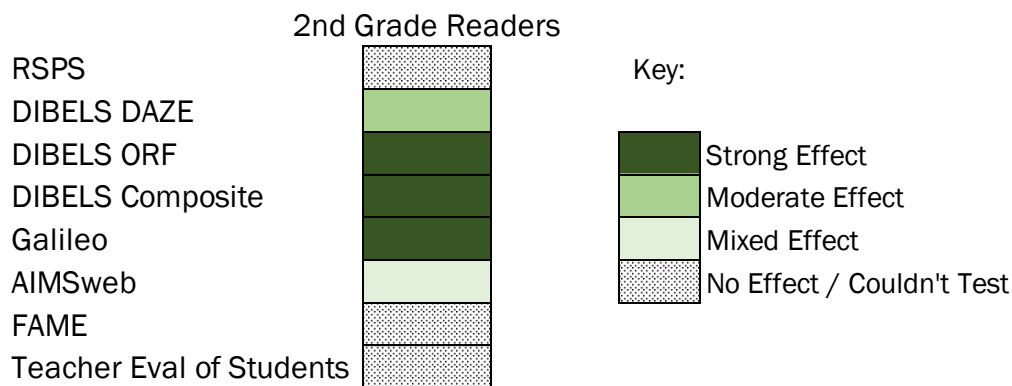
**Figure 33: Teacher Evaluations of 2nd Grade Participants – Unpaired Assessments**

	Pre Test	Post Test	Score Increase
Concentration	3	3.77	0.77
Enjoyment	3.27	3.85	0.58
Comprehension	3.13	3.23	0.1
Attendance	4	3.89	-0.11

~~

Overall, the examined assessments provide varying evidence of RBBB program effectiveness for 2<sup>nd</sup> grade Readers. The program appears to be much more effective at the 3<sup>rd</sup> grade level. However, these results may be due to the smaller sample size of the pilot program. As well, the pilot was administered in only three school districts, which may have unique impacts on 2<sup>nd</sup> grade reading achievements that aren't tempered by a larger sample of districts like the 3<sup>rd</sup> grade Readers have. The following figure visualizes the strength of evidence of the 2<sup>nd</sup> grade pilot program for RBBB Readers.

**Figure 34: Summary of Data Results for 2<sup>nd</sup> Grade Pilot Program, by Assessment**



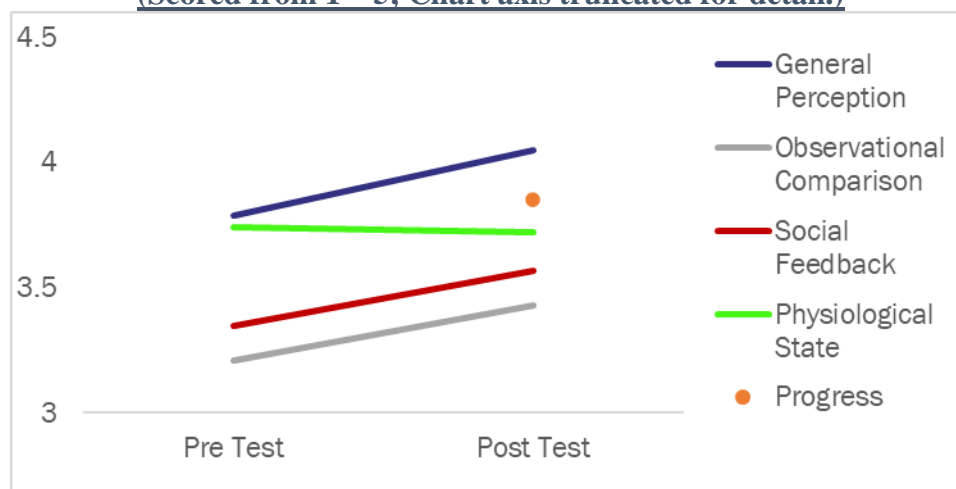
## Results – School Day Pilot Program

RBBB conducted another pilot recently, implementing the program during the school day instead of after school time at Longview Elementary School in the Osborn school district. This section discusses those results.

### *Reader Self-Perception Scale (“RSPS”)*

Figure 35 below shows results of School Day pilot program participants’ changes in scores on the RSPS. These students did not show statistically significant changes in their RSPS scores from pre-test to post-test time, although they do show increases. A small sample size (n=20) likely explains this.

**Figure 35: RSPS Subscale Component Score Averages for 3rd Grade Participants**  
**(Scored from 1 – 5; Chart axis truncated for detail.)**



	Pre Test	Post Test	Score Increase
General Perception	3.79	4.05	0.26
Progress	N/A	3.85	N/A
Observational Comparison	3.21	3.43	0.22
Social Feedback	3.35	3.57	0.22
Physiological State	3.74	3.72	-0.02

\* Paired samples t-tests were conducted with significance set at the  $\alpha = .05$  level.

This analysis also explored the potential impact that attendance in RBBB sessions could have on progress. For the RSPS, on average, the number of RBBB program sessions attended did not notably correlate with greater increases on the RSPS subscale scores.<sup>19</sup>

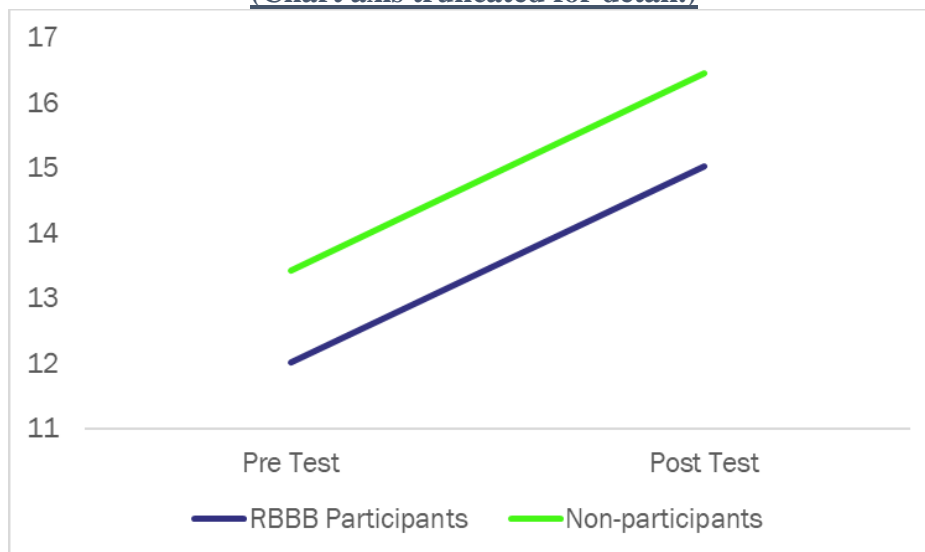
<sup>19</sup> Bivariate Pearson correlation with statistical significance measured at the  $\alpha = .05$  level.

## ORA

ORA score data include both RBBB program participants and non-participants' scores, and applied a matched-comparison group approach with GLM modeling as described previously.

For Readers in the RBBB program, their ORA score increased by exactly 3 points from pre-test to post-test time, while the comparison group's score increased almost the same amount. The following visualization and table show these results.

**Figure 36: ORA Average Scores, Pre and Post Test**  
(Chart axis truncated for detail.)



	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	12.02	15.02	3	n=23
Non-participants	13.43	16.45	3.02	n=29

## NWEA MAP Test – RIT

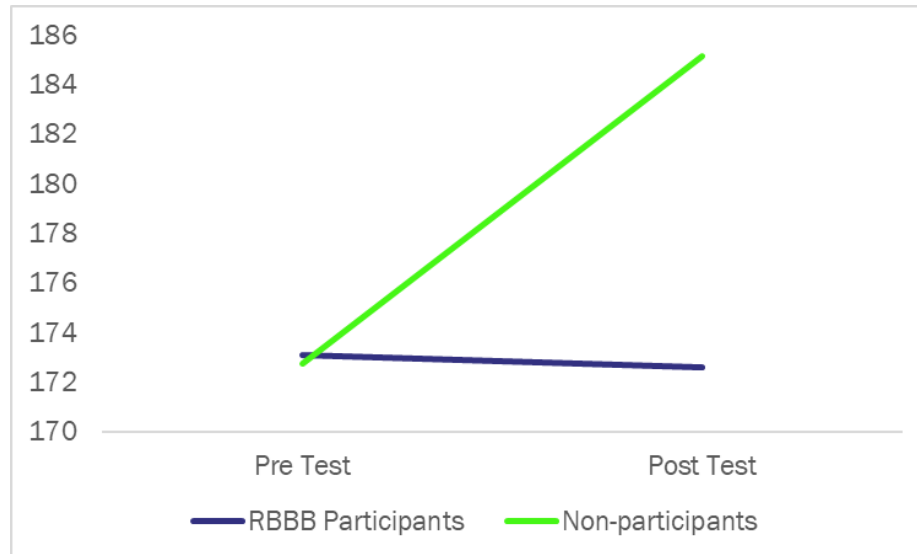
The Osborn school district implements the MAP test and reports RIT scores for all examined semesters.

RIT data include both RBBB program participants and non-participants' scores. These scores can be compared with published RIT Reading Student Achievement Norms, however, the sample size in the School Day pilot is too small to separate into Fall / Spring analyses. Therefore, both semesters are analyzed together and not compared to the benchmark norms.

This analysis applies a matched-comparison group approach when analyzing the effect of RBBB School Day pilot program participation on RIT scores, with GLM modeling, as described before.

For Readers in the RBBB School Day pilot program, their RIT scores slightly decreased, on average, while their comparison group members' scores increased by over 12 points. The following figures show these results.

**Figure 37: RIT Average Scores, Pre and Post Test, Fall Semesters only**



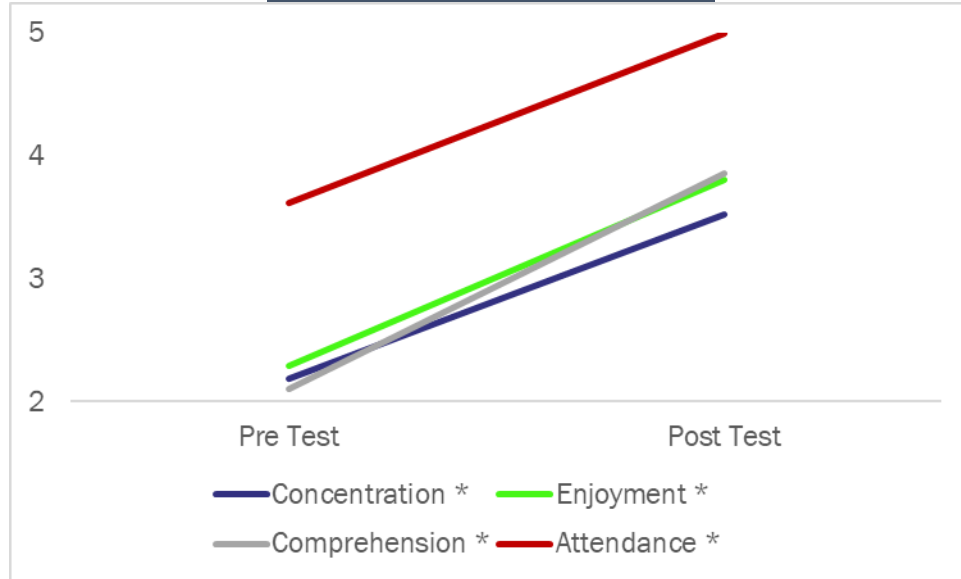
	Pre Test	Post Test	Score Increase	Group Size
RBBB Participants	173.1	172.61	-0.49	n=31
Non-participants	172.77	185.15	12.38	n=100

In the School Day pilot, RBBB participants started at pre-test time with a slightly higher average RIT score than their matched comparison group but didn't gain. This may be explained by a small sample size.

#### *Teacher Evaluation of Students – “Reader Teacher Survey”*

Teachers evaluated their 3<sup>rd</sup> grade Readers who participated in the RBBB program School Day pilot in four areas – Focus, Enjoyment of Reading, Reading Comprehension, and Attendance. Figure 21 below shows the average score for participants from the beginning of the semester (“pre”) to the end of the semester (“post”). On average, participants improved on all four subscales to a statistically significant extent. The subscales were scored from 1 to 5, with a higher score denoting greater teacher agreement with the student's achievement. The following visualization and table show these results.

***Figure 38: Teacher Evaluations of 3rd Grade Participants***  
***(Chart axis truncated for detail.)***



	Pre Test	Post Test	Score Increase
Concentration	2.19	3.52	1.33 *
Enjoyment	2.29	3.81	1.52 *
Comprehension	2.1	3.86	1.76 *
Attendance	3.62	5	1.38 *

*\* Denotes a statistically significant improvement from pre- to post-time measures, at the  $\alpha = .05$  level, according to paired-samples t-test.*

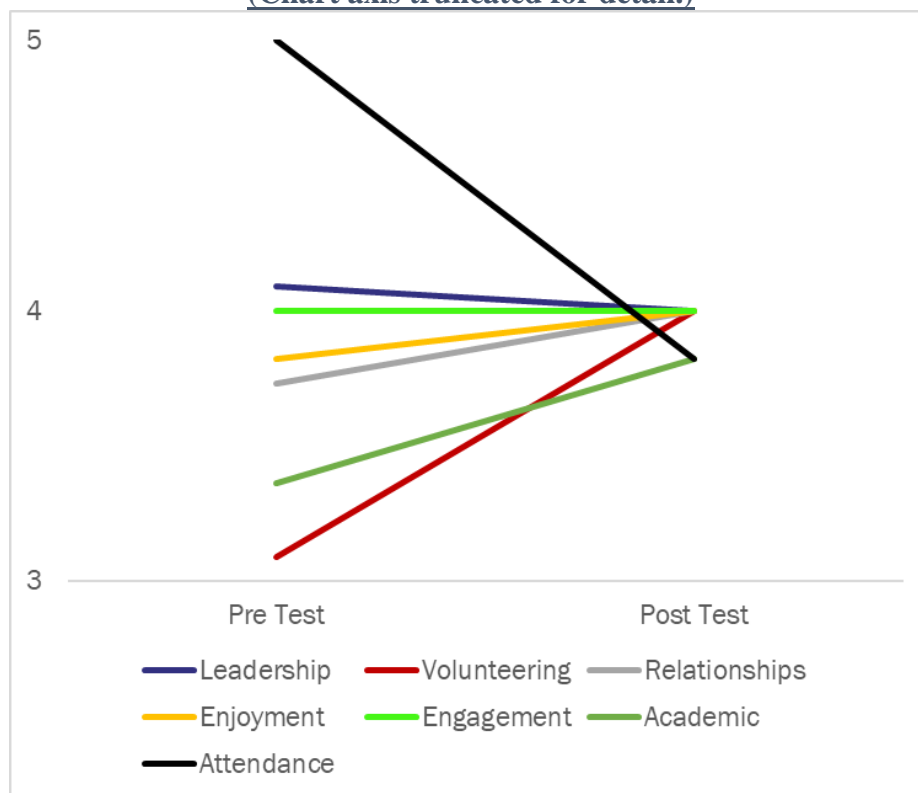
Beyond the teachers' perception of attendance, the data allow for testing correlation between attendance and increases in concentration, enjoyment, and comprehension. On average, attendance was positively correlated with greater increases on the teacher evaluation scores. The more sessions a student attended, the higher their increase in teacher evaluation score from pre-test to post-test time.

#### *Teacher Evaluation of Leaders – “Leader Teacher Survey”*

Teachers evaluated the 6<sup>th</sup> grade Leaders who participated in the RBBB School Day pilot program in seven areas – Leadership skills, Volunteer initiative, Positive relationships, Enjoyment of school, Engagement in the classroom, Academic expectations, and Attendance. Figure xx below shows the average score for participants from the beginning of the semester (“pre”) to the end of the semester (“post”). Participants improved on 4 of the 7 subscales, and statistically significantly improved on the subscale measuring their Volunteer initiative. The

subscales were scored from 1 to 5, with a higher score denoting greater teacher agreement with the student's achievement. The following visualization and table show these results.

***Figure 39: Teacher Evaluations of 6th Grade Participants***  
(Chart axis truncated for detail.)



	Pre Test	Post Test	Score Increase
Leadership	4.09	4	-0.09
Volunteering	3.09	4	0.91 *
Relationships	3.73	4	0.27
Enjoyment	3.82	4	0.18
Engagement	4	4	0
Academic	3.36	3.82	0.46
Attendance	5	3.82	-1.18

*\* Denotes a statistically significant improvement from pre- to post-time measures, at the  $\alpha = .05$  level, according to paired-samples t-test.*

For the RBBB Leaders' teacher evaluation data, a completed pre and post-test must be in the data for each student in order to examine their improvement. This was the case in the data for 11 students, a small sample.

Beyond the teachers' perception of attendance, the data allow for testing for correlations between attendance and increases in leadership, volunteering, relationships, enjoyment, engagement, and



academic attainment. No informative correlations emerge between Leaders' attendance and changes in their teachers' perceptions of them.

### *Reading Leader Survey*

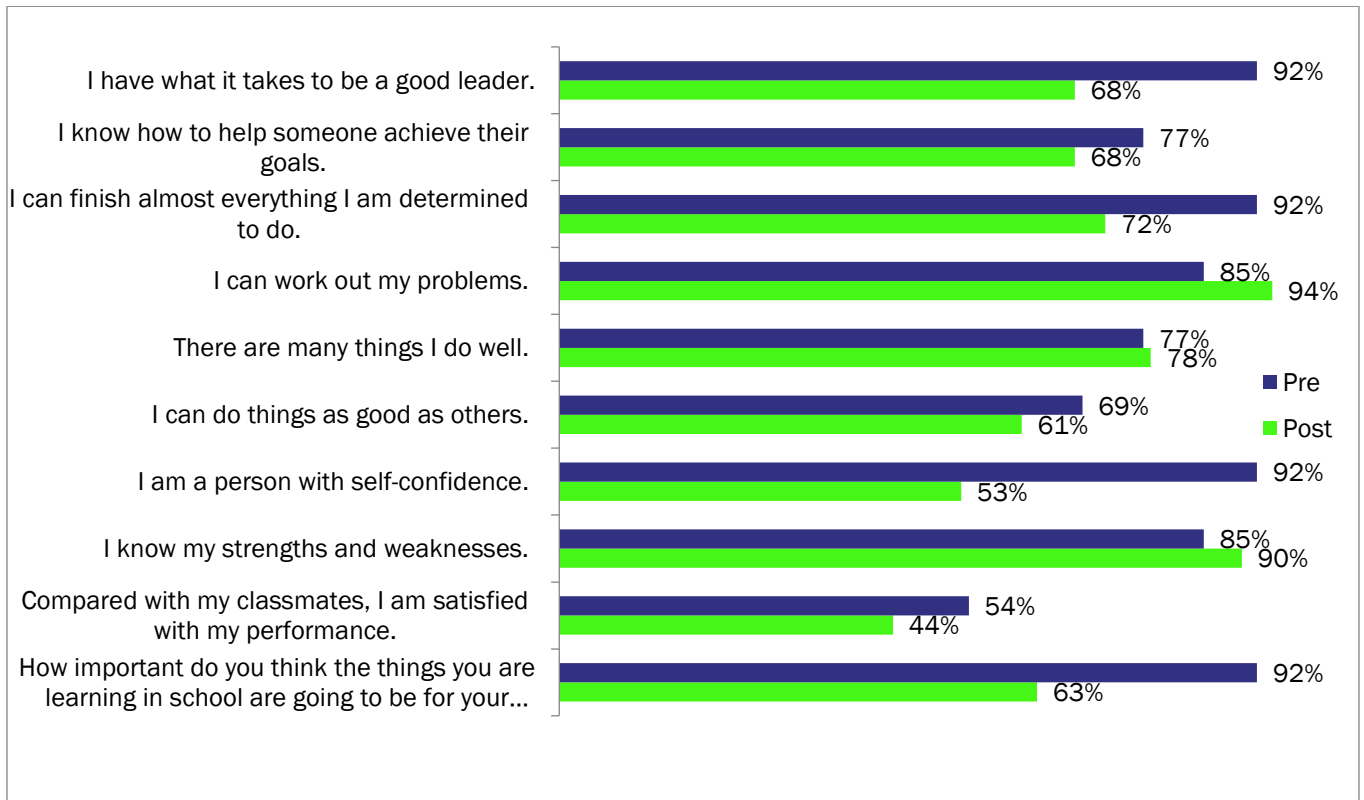
Not all schools have pretest data, and not all schools have post-test data. Like the teacher evaluation data, a completed pre-test and post-test must be present in order to assess student improvement. This was the case in the data for 235 students.

For these students, their average post-test time score was similar, and slightly decreased, from their average pre-test time score. Questions on the Reading Leader Survey are scaled from 1 to 5, with 5 indicating greater agreement with the positive statements. Overall, these 6<sup>th</sup> grade Leaders in the School Day pilot program went from an average score of **4.3** to **3.9** throughout their time in the program.

The data also allow a specific look at each question to see whether there are large variations in student response from pre-test to post-test time. After examining all of the questions, all of the questions on the survey showed either **no change or a decrease** in student responses. None of the questions showed statistically significant improvement from pre-test to post-test time. However, within the available data, only 9 students took both a pre-test and a post-test. These 9 students differ from the averages shown in the figure below.

Figure xx below reports the percentage of students responding that they either "Strongly Agree" or "Agree" with each statement.

**Figure 40: Percentage of 6<sup>th</sup> Grade Leaders either Strongly Agreeing or Agreeing with each Statement, RLS**



~~

Overall, little evidence exists that the Readers or Leaders in the School Day pilot improved on the provided assessments as a result of participating in the RBBB program.

## Matched Leader Attendance

One final question was queried of the data – whether matched Leader attendance impacts Readers’ achievements. In other words, when a Reader has the same Leader for more RBBB program sessions, does this correlate with their reading achievements?

FirstEval examined this by merging session attendance data by Reader -Leader pair with the assessment data. These data were available from the Avondale school district from the Fall, 2018 semester and from the Buckeye school district from the Spring 2019 semester. This section discusses those results by assessment.

When measuring achievement via the **RSPS** for both Avondale and Buckeye, no correlation was found between Leaders’ number of sessions attended and their Readers’ achievements. (n=72 matched pairs)

When measuring achievement via the **DIBELS DAZE** for both Avondale and Buckeye, no correlation was found between Leaders’ number of sessions attended and their Readers’ achievements. (n=105 matched pairs)

When measuring achievement via the **DIBELS ORF** for both Avondale and Buckeye, a small positive correlation was found between Leaders’ number of sessions attended and their Readers’ achievements that neared statistical significance.<sup>20</sup> (p=.088) (n=105 matched pairs)

When measuring achievement via the **DIBELS Composite** for the Buckeye school district, a small positive correlation was found between Leaders’ number of sessions attended and their Readers’ achievements. (n=54 matched pairs)

When measuring achievement via the **Galileo** assessment for the Avondale school district, no correlation was found between Leaders’ number of sessions attended and their Readers’ achievements. (n=51 matched pairs)

For the **FAME** assessment in the Avondale school district, there weren’t enough matched pairs to examine their impact on Reader movement between Partially Proficient to Proficient and between Proficient to Highly Proficient.

When measuring achievement via the **Teacher Evaluation of Students** for both Avondale and Buckeye, no correlation was found between Leaders’ number of sessions attended and their Readers’ increases in concentration, enjoyment, or comprehension. (n=63 matched pairs)

In sum, little evidence was found that a matched Leader contributes to Reader achievement.

---

<sup>20</sup> Bivariate Pearson correlation with statistical significance measured at the  $\alpha=.05$  level.

## Summary

The RBBB program has notable effects on its participants' literacy skills. Most importantly, for 3<sup>rd</sup> grade participants, or "Readers," program participation has strong effects on their social feedback skills, physiological state, and their rating score of general perception. Importantly, 3<sup>rd</sup> grade participants show significant improvement in reading according to the RSPS assessment, DIBELS Composite and ORF assessments, Galileo assessment, ORA scores, and AIMSweb scores. And, teacher evaluations' of Readers show significant improvement for RBBB program participants and some significant improvements for 8<sup>th</sup> grade Leaders.

For Leaders, program participation shows no or moderate effects. This may be because Leaders selected for program participation are high performers, and any progress in their reading levels or social and personal responsibility would be minimal in one semester.

When available, attendance information was incorporated into this report's analyses. Increased attendance often showed a correlation with stronger assessment results. This trend of overall positive correlations is another indication of program effectiveness.

As well, assessments provide varying to moderate evidence that the 2<sup>nd</sup> grade pilot program for Readers effects participants' literacy skills. While evidence for the 2<sup>nd</sup> grade pilot is positive, it is not as strong as findings for the 3<sup>rd</sup> grade Readers.

For RBBB participants in the School Day pilot, there was little evidence that their program participation was impactful. This may be due to a small sample size.

RBBB has complex program data, in that student progress is tracked over multiple years and locations with multiple assessment instruments. Data richness is increased with the availability of assessment score data form a comparison group. This report reflects the examination of the RBBB data and the comparison data. The numerous assessment instruments that the RBBB program employs is commendable, showing that RBBB truly aims to measure its program participant progress through many angles and in cooperation with participating schools.

As the third evaluation report of the RBBB program, this allows for reflection on program growth and achievement. The first report, provided in 2016, showed RBBB effectiveness but had small sample sizes, as the program was still small. The 2018 report and this report benefit from larger sample sizes as more and more schools have joined the program. The 2018 report, and this one, also show RBBB effectiveness but are bolstered by increasing sample sizes as more and more schools are added to the program. This strengthens the findings that the program is having an impact on 3<sup>rd</sup> grade readers.

Overall, the RBBB program is significantly improving reading for its 3<sup>rd</sup> grade participants, and provides evidence of improving literacy for 2<sup>nd</sup> grade students, as evidenced by numerous assessment instruments.

## Appendix A – RBBB Logic Model with Report Scope Highlighted in Red

### Target Population:

- 3<sup>rd</sup> grade
- Tier II
- Fluency >60wpm
- Not receiving additional intervention



Read Better Be Better

### Read Better Be Better Logic Model

The red-shaded area represents the data examined for this report.

Inputs	Activities	Outputs	Short	Medium	Long
Staff expertise and experience	Proprietary reading comprehension curriculum.	# Students received program	Attendance – RBBB and school (RBBB log and school attendance data with grade average as control)	Improved reading comprehension (Galileo testing with grade average as control)	Better learners (Approaches to Learning)
Research	o 2x/week	# Student Hours	Retention (RBBB log)	Improved concentration (3 <sup>rd</sup> grade teacher pre-post evaluation)	Increased high school graduation rates
Collaborative Partnerships	o Minimum one semester	% Students completing program	Increased enjoyment of reading (3 <sup>rd</sup> grade teacher pre-post evaluation)	Increased sense of personal responsibility (Social & Personal Responsibility Scale and pre-post self-evaluation)	(longitudinal study of school graduation data)
Time	8 <sup>th</sup> grade “reading leaders” trained to implement curriculum	# 8 <sup>th</sup> grade trained “reading leaders”	Enjoyment of RBBB program (post self-evaluation)	Improved general academic performance (longitudinal study of school Math and ELA scores)	
Funds	Be A Better Reader		Increased self-efficacy (pre-post self-evaluation: The Reader Self-Perception Scale)		
	Be A Better Thinker				
	Be A Better Mover				
<b>Assumptions</b>			<b>External Factors</b>		
Achieving a reading age of 10 allows independent learning and positively impacts earning potential			Current poor AZ literacy stats < MOWR law < AZ Literacy Plan. Current high school drop-out rates. Rise of service-learning model.		

## Appendix B: Reader Self-Perception Scale (“RSPS”)

**Instruction:** Below are statements about reading. Please read each statement carefully. Then fill in the bubbles that show how much you agree or disagree with the statement.

Example:

If you are really positive that pepperoni pizza is best, fill in the bubble under “**Strongly Agree**”.

If you think that is good but maybe not great, fill in the bubble under “**Agree**”.

If you can't decide whether or not it is best, fill in the bubble under “**Undecided**”.

If you think that pepperoni pizza is not all that good, fill in the bubble under “**Disagree**”.

If you are really positive that pepperoni pizza is not very good, fill in the bubble under “**Strongly Disagree**”.

Now, please fill in the bubbles that show how much you agree or disagree with each of the following statement.

1. I think I am a good reader..... ☐ ☐ ☐ ☐ ☐
2. I can tell that my teacher likes to listen to me read..... ☐ ☐ ☐ ☐ ☐
3. I read faster than other kids..... ☐ ☐ ☐ ☐ ☐
4. My teacher thinks that my reading is fine..... ☐ ☐ ☐ ☐ ☐
5. I like to read aloud..... ☐ ☐ ☐ ☐ ☐
6. When I read, I can figure out words better than other kids.... ☐ ☐ ☐ ☐ ☐
7. My classmates like to listen to me read..... ☐ ☐ ☐ ☐ ☐
8. I feel good inside when I read..... ☐ ☐ ☐ ☐ ☐
9. My classmates think that I read pretty well..... ☐ ☐ ☐ ☐ ☐
10. When I read, I don't have to try as hard as I used to..... ☐ ☐ ☐ ☐ ☐
11. I seem to know more words than other kids when I read. ☐ ☐ ☐ ☐ ☐
12. People in my family think I am a good reader..... ☐ ☐ ☐ ☐ ☐
13. I am getting better at reading..... ☐ ☐ ☐ ☐ ☐
14. I understand what I read as well as other kids do..... ☐ ☐ ☐ ☐ ☐
15. When I read, I need less help than I used to..... ☐ ☐ ☐ ☐ ☐
16. Reading makes me feel happy inside..... ☐ ☐ ☐ ☐ ☐
17. My teacher thinks I am a good reader..... ☐ ☐ ☐ ☐ ☐
18. Reading is easier for me than it used to be..... ☐ ☐ ☐ ☐ ☐
19. I read faster than I could before..... ☐ ☐ ☐ ☐ ☐
20. I read better than other kids in my class..... ☐ ☐ ☐ ☐ ☐
21. I feel calm when I read..... ☐ ☐ ☐ ☐ ☐
22. I read more than other kids..... ☐ ☐ ☐ ☐ ☐
23. I understand what I read better than I could before..... ☐ ☐ ☐ ☐ ☐
24. I can figure out words better than I could before..... ☐ ☐ ☐ ☐ ☐
25. I feel comfortable when I read..... ☐ ☐ ☐ ☐ ☐
26. I think reading is relaxing..... ☐ ☐ ☐ ☐ ☐
27. I read better now than I could before..... ☐ ☐ ☐ ☐ ☐
28. When I read, I recognize more words than I used to..... ☐ ☐ ☐ ☐ ☐
29. Reading makes me feel good..... ☐ ☐ ☐ ☐ ☐
30. Other kids think I'm a good reader..... ☐ ☐ ☐ ☐ ☐
31. People in my family think I read pretty well..... ☐ ☐ ☐ ☐ ☐
32. I enjoy reading..... ☐ ☐ ☐ ☐ ☐
33. People in my family like to listen to me read..... ☐ ☐ ☐ ☐ ☐

## Appendix C: Reader Teacher Survey

School:

Teacher Name:



Read Better Be Better

### Reader Teacher Survey

Student Name:

Student ID #:

Grade:

Thank you for taking the time to fill out this survey. This is one of our main tools for evaluating the success of our program and your feedback is very important and greatly appreciated. If you have any questions or would like to see the report at the end of the semester please contact your Read Better Be Better District Leader. Again thank you so much for taking the time to fill out this survey and supporting Read Better Be Better.

#### Pre - Evaluation

Please fill out this section at the BEGINNING of program.

This student is able to focus on reading material from the beginning to the end

☐ Strongly Disagree
 ☐ Disagree
 ☐ Neutral
 ☐ Agree
 ☐ Strongly Agree

This student appears to enjoy reading self-selected books

☐ Strongly Disagree
 ☐ Disagree
 ☐ Neutral
 ☐ Agree
 ☐ Strongly Agree

This student understands reading material in class

☐ Strongly Disagree
 ☐ Disagree
 ☐ Neutral
 ☐ Agree
 ☐ Strongly Agree

This student attends class on a regular basis

☐ Strongly Disagree
 ☐ Disagree
 ☐ Neutral
 ☐ Agree
 ☐ Strongly Agree

Additional Comments:

#### Post - Evaluation

Please fill out this section at the END of program.

This student is able to focus on reading material from the beginning to the end

☒ Strongly Disagree
 ☒ Disagree
 ☒ Neutral
 ☒ Agree
 ☒ Strongly Agree

This student appears to enjoy reading self-selected books

☒ Strongly Disagree
 ☒ Disagree
 ☒ Neutral
 ☒ Agree
 ☒ Strongly Agree

This student understands reading material in class

☒ Strongly Disagree
 ☒ Disagree
 ☒ Neutral
 ☒ Agree
 ☒ Strongly Agree

This student attends class on a regular basis

☒ Strongly Disagree
 ☒ Disagree
 ☒ Neutral
 ☒ Agree
 ☒ Strongly Agree

Additional Comments:

www.readbetterbebetter.org

Updated: 2.11.2020\SF

## Appendix D: Leader Teacher Survey

School:

Teacher Name:



### Read Better Be Better Leader Teacher Survey

Student Name:

Student ID #:

Grade:

Thank you for taking the time to fill out this survey. This is one of our main tools for evaluating the success of our program and your feedback is very important and greatly appreciated. If you have any questions or would like to see the report at the end of the semester please contact your Read Better Be Better District Leader. Again thank you so much for taking the time to fill out this survey and supporting Read Better Be Better.

#### Pre - Evaluation

Please fill out this section at the BEGINNING of program.

This student displays strong leadership skills

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

This student has taken opportunities to volunteer and help others

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

This student has positive relationships with younger students in their school

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

This student seems to enjoy school

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

This student is engaged in the classroom

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

This student is meeting or exceeding academic expectations

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

This student attends class on a regular basis

☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree

Additional Comments:

#### Post - Evaluation

Please fill out this section at the END of program.

This student displays strong leadership skills

☒ Strongly Disagree ☒ Disagree ☒ Neutral ☒ Agree ☒ Strongly Agree

This student has taken opportunities to volunteer and help others

☒ Strongly Disagree ☒ Disagree ☒ Neutral ☒ Agree ☒ Strongly Agree

This student has positive relationships with younger students in their school

☒ Strongly Disagree ☒ Disagree ☒ Neutral ☒ Agree ☒ Strongly Agree

This student seems to enjoy school

☒ Strongly Disagree ☒ Disagree ☒ Neutral ☒ Agree ☒ Strongly Agree

This student is engaged in the classroom

☒ Strongly Disagree ☒ Disagree ☒ Neutral ☒ Agree ☒ Strongly Agree

This student has met or exceeds academic expectations

☒ Strongly Disagree ☒ Disagree ☒ Neutral ☒ Agree ☒ Strongly Agree

This student attends class on a regular basis

☒ Strongly Disagree ☒ Disagree ☒ Neutral ☒ Agree ☒ Strongly Agree

Additional Comments:

[www.readbetterbebetter.org](http://www.readbetterbebetter.org)

Updated: 2.11.2020\SF



## Appendix E: Reading Leader Survey



### Reading Leader Survey - POST

Read each statement. Fill in the bubble that best describes how much you agree or disagree with the statement about yourself.

Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. I have what it takes to be a good leader.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. I know how to help someone achieve their goals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. I can finish almost everything I am determined to do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. I can work out my problems.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. There are many things I do well.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. I can do things as good as others.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. I am a person with self-confidence.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. I know my strengths and weaknesses.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Compared with my classmates, I am satisfied with my performance.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. How important do you think the things you are learning in school are going to be for your later life?	Very Important <input type="radio"/>	Quite Important <input type="radio"/>	Fairly Important <input type="radio"/>	Slightly Important <input type="radio"/>	Not at all Important <input type="radio"/>
11. In the past 10 weeks (since RBBB started) how often did you:	Almost Always	Often	Sometimes	Seldom	Never
Enjoy being in school?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hate being in school?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Try to do your best work in school?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Appendix F – Detailed Statistical Methodology

The rich data in this project provided an opportunity to isolate the effects of the RBBB program because it often contained large comparison groups. This allowed for the construction of a matched comparison group. As well, the data contain repeated measures (i.e., the DIBELS ORF and ORA scores measured at pre-test and post-test time), matched at the student level.

Hierarchical linear modeling (“HLM”) is a popular design model in educational evaluation. HLM considers students to be “nested” within schools, as a way of controlling for potential school-level effects while isolating program effects on students. While our overall dataset was large ( $n > 1,000$ ), the multiple school locations resulted in small sub-samples. These sub-sample sizes were not large enough to analyze the data at a nested level, and therefore it was deemed ineligible for HLM.

The data were also explored for their eligibility for propensity score matching methods (“PSM”). Precedence exists for conducting PSM instead of HLM in educational evaluation settings.<sup>21</sup> PSM mimics randomization in a quasi-experimental setting by matching treated results (i.e., the RBBB program participants’ DIBELS ORF and ORA score changes) with untreated results (the non-participants’ DIBELS ORF and ORA score changes). In order to conduct PSM on a dataset, certain assumptions must be met (conditional independence, common support, and stable unit treatment values). As well, the beauty of using PSM is the ability to control for varying factors and isolate the treatment effect. In our data, eligible control factors among both the treated and untreated groups are gender, race/ethnicity, and free and reduced lunch status. RBBB participants don’t significantly differ on gender make-up compared to non-participants at any grade level. Sometimes RBBB participants differ on race/ethnicity and free and reduced lunch status from non-participants. Given that these differences were not stark, and that the data did not pass all of the PSM assumptions, it wasn’t deemed eligible for PSM methods.

For these reasons, the best approach was to employ general linear modeling (“GLM”). GLM provides a way to test whether the change in a repeated measure (i.e. DIBELS ORF and ORA scores pre and post-test time) differs between a treated group (RBBB program participants) and an untreated group (non-participants).

For most assessments, RBBB provided FirstEval with comparison group data. With this, FirstEval was able to construct a comparison group that closely matched the pre-test scores of the participants. In using this matched comparison group approach, the RBBB participants and the non-participants had similar starting points. The original comparison group datasets were culled down to be within +/- 1.0 standard deviations of the participants’ pre-test mean scores. This GLM matched comparison group approach is reported in this document.

---

<sup>21</sup> See Lane et al., 2012, An Illustrative Example of Propensity Score Matching with Education Research, [https://www.researchgate.net/publication/273061804\\_An\\_Illustrative\\_Example\\_of\\_Propensity\\_Score\\_Matching\\_with\\_Education\\_Research](https://www.researchgate.net/publication/273061804_An_Illustrative_Example_of_Propensity_Score_Matching_with_Education_Research)

## Appendix G – Comparing Results with Published Literature

RBBB’s signature program can be compared to a program called Experience Corps (“EC”), where older adult volunteers are placed in elementary schools to tutor elementary school students in reading.

Two published studies of the Experience Corps program, Lee et al. and Rebok et al.<sup>22</sup>, show positive program effects of volunteering on the elementary school students’ literacy skills gains. The Rebok et al. paper used the PPVT-III and CTBS assessments, while the Lee et al. paper used the PPVT and WJ-PC assessments. No published papers used the same assessments that RBBB measures.

Overall, the RBBB program shows stronger results than the EC program in fluency and appears to be similar with the EC program for reading comprehension. With regards to reading comprehension, both the RBBB program and the EC program had positive, but overall mixed results.

The figure below details the findings from both programs:

	RBBB Evaluation - 3rd graders	Rebok	Lee
<u>Measurement</u>			
Teacher Perception	Teacher Survey: Significant increase in concentration, enjoyment, and comprehension for participants	N/A	"97% of teachers find EC beneficial to the participants" (not statistically tested)
Oral Reading Fluency	Dibels ORF: Statistically significant, large n positive impact on RBBB participants. AIMSweb ORF: Statistically significant gain for participants compared to non-participants	PPVT-III: First grade participants; 3.13% score increase for participants, statistically significant over the control	PPVT: No statistically significant impact shown here
Comprehension	Dibels DAZE: Statistically significant increase for participants, but gains are less than the comparison groups' gains AIMSweb Comp: Statistically significant gain for participants compared to non-participants.	CTBS: 2/3 of EC schools have growth over the control school, but not statistically significant	WJ-PC: Participants' gains are statistically significantly greater than control group with a 1.5 effect size.

<sup>22</sup> “The Effect of the Experience Corps Program on Student Reading outcomes” by Lee, Yung Soo et al. 2010. In *Education and Urban Society*; and “Short Term Impact of Experience Corps Participation on Children and Schools: Results from a Pilot Randomized Trial” by Rebok, George W. et al. 2004. In *Journal of Urban Health: Bulletin of the New York Academy of Medicine*.